12-2010

Incorporating Spatial Information into Gas Plume Detection in Hyperspectral Imagery

Cameron S. Grant
Utah State University

INCORPORATING SPATIAL INFORMATION INTO GAS PLUME DETECTION IN HYPERSPECTRAL IMAGERY

by

Cameron S. Grant

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Electrical Engineering

Approved:

Dr. Jacob H. Gunther Major Professor	Dr. Todd K. Moon Committee Member
Dr. Bedri Cetiner Committee Member	Dr. Byron R. Burnham Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2010

Abstract

Incorporating Spatial Information into Gas Plume Detection in Hyperspectral Imagery

by

Cameron S. Grant, Master of Science

Utah State University, 2010

Major Professor: Dr. Jacob H. Gunther
Department: Electrical and Computer Engineering

Detection of chemical plumes in hyperpsectral data is a problem having solutions that focus on spectral information. These solutions neglect the presence of the spatial information in the scene. The spatial information is exploited in this work by assignment of prior probabilities to neighborhood configurations of signal presence or absence. These probabilities are leveraged in a total probability approach to testing for signal presence in a pixel of interest. The two new algorithms developed are named spatial information detection enhancement (SIDE) and bolt–on SIDE (B–SIDE).

The results are explored in comparison to the clutter matched filter (CMF), a standard spectral technique, and to several supervised machine learning techniques. The results show a great improvement of SIDE over these other techniques, in some cases showing the poorest performance of the SIDE filter being much better than the CMF at its best.

(130 pages)

To my sweetheart Kirsten

Acknowledgments

I would like to thank Dr. Jacob Gunther and Dr. Todd Moon for their wise guidance and patient offerings of insight, Matthew Stites for his willingness to provide those same things, and my wife, Kirsten, for the constant support and encouragement.

Cameron S. Grant

Contents

List of Tables

List of Figures

Acronyms

DIRS	Digital Imaging and Remote Sensing (Laboratory)
DIRSIG	DIRS Image Generation Model
LWIR	Long–Wave Infrared
SWIR	Short–Wave Infrared
VIS	Visible
ACE	Adaptive Coherence Estimator
CMF	Clutter Matched Filter
SIDE	Spatial Information Detection Enhancement
B–SIDE	Bolt–On Spatial Information Detection Enhancement
LR	Likelihood Ratio
LLR	Log-Likelihood Ratio
HSI	Hyperspectral Imaging
LDA	Linear Discriminant Analysis
QDA	Quadratic Discriminant Analysis
LC	Least–Squares Linear Classifier
SVM	Support Vector Machine
MMSE	Minimum Mean Square Error

Chapter 1

Introduction

The aim of this thesis is to explore the problem of pattern recognition of spatially distributed objects with amorphous shapes (no definite form) and potentially changing shapes. Such objects may arise, for example, from phenomena such as gas plume emissions from an industrial stack or a vehicle exhaust pipe, the spread of a plant species across the countryside, the dispersion of oil from an oil spill, or the infectious spread of a disease. It is assumed that there is some measurable but weak signature vector at each collection point, and that the collection points (pixels) are in a regular spatial array or representative image, enabling detection. By exploiting spatial or geometric information inherent in the phenomenon, it is possible to improve detection (lower detection limits) and reduce false detections. In many other pattern recognition tasks, spatial and geometric information has been used by identifying particular shapes (e.g., airplanes) or utilizing particular features (e.g., straight or abrupt edges). However, this work focuses on objects that have no well-defined geometric features. For example, a gas plume may meander in variable winds so that the edges of the plume may not be straight or the plume could detach into separate, but still locally contiguous puffs. Despite this variability, there is information that may be exploited by assuming spatial contiguity, a physically-based assumption.

It is assumed that the data have sufficient resolution that the object of interest covers at least several pixels in the image and that the phenomenon exhibits a fairly uniform "spreading" behavior; it spreads from one or more points forming a distribution with spatial continuity. Under these assumptions two conditions are likely for a given pixel neighborhood: either a pixel is similar to its neighbors (either all neighbors "on" or all "off") or a continuous, but potentially curvy boundary exists which divides two contiguous regions. One might expect that a neighborhood of randomly distributed "on" and "off" pixels would

be unlikely. In summary, it is assumed that spatial contiguity occurs more often than spatial irregularity.

This work develops a detection-theoretic framework for modeling and exploiting spatial contiguity for pattern recognition. This work also establishes a probabilistic *a priori* model that can quantify spatial contiguity and irregularity. This model assigns to a pixel a value which represents the probability of the signal presence based on neighborhood pixels and their physical configuration. While exhaustive physical modeling could be used to exactly determine the probability of each physical neighborhood configuration, it has been found more useful (and more tractable) to represent the probabilities with a low-complexity one-parameter spatial probability model.

This work uses the prior spatial probability model to compute a posterior probability for a pixel in a *maximum a posteriori* (MAP) setting. It is shown that the resulting detector can be used as a spatial filter to augment any chosen pixel-by-pixel detection scheme for detecting the presence of signal at each pixel. This work uses the pixel-by-pixel detection scheme output as the likelihood function for the posterior model. This neighborhood processing model, called Spatial Information Detection Enhancement (SIDE), supplements and improves existing exploitation techhologies and techniques.

SIDE is applied to the problem of plume detection in hyperspectral imagery to demonstrate its performance and utility. To quantify the detector performance, data was generated using the industry-standard DIRSIG program [1] to provide known scenes (or ground truth) for experiments. This work demonstrates with these experiments that the use of SIDE for post-processing obtains comparable receiver operating characteristic (ROC) curves on a plume ten-times less dense to the ROC generated by traditional processing alone — a factor of 10x increase in performance by this measure.

1.1 Related Work

Image object and pattern recognition are diverse fields wherein a number of different approaches have been developed for recognition of patterns, shapes, objects, and other entities of interest. One prominent method, the scale-invariant feature transform (SIFT),

uses extracted key points of objects in an image to provide a feature description of the object [2]. This object description is invariant with respect to scale, orientation, and affine transformations and is robust with respect to illumination changes and noise. Many SIFT variations which use extracted key points for object description and detection exist [3] and have successfully been used for rigid and semi-rigid objects such as bicycles, airplanes, and human faces in a variety of environments, even with partial occlusion and changing object orientation with respect to the camera. There are statistical methods that address the variance in object features [4] that have been applied to the problem of recognizing object classes, which may vary in specific appearance or shape [5]. However, these methods require that the objects have definable spatial relationships among features such as a predictable set of edge relationships.

Where the object has no predefined shape, pattern recognition techniques are relied on more heavily. One example that combines object recognition approaches with statistical pattern recognition is detecting sports players [6]. These approaches take advantage of a separate feature, such as spectral information (e.g., color), to extend other object recognition techniques.

Gas plume detection, identification, and characterization using exploitation of hyperspectral image data has focused on strictly spectral techniques and has been reasonably successful. Foy provides a succinct account of current approaches [7]. Researchers have published a number of comparisons that show a matched filter using a generalized least squares (GLS) approach is most successful [8–10]. The GLS approach has been refined [11–13] with various methods used to characterize and subtract the background clutter [14,15]. Research in this field includes independent component analysis [16], sparse matrix techniques [17], and a method using endmember extraction to distinguish background clutter from plume [18].

Some attempts have been made to include spatial information in material identification including a simplex optimization technique [19], and using support vector machines (SVM) to emphasize contiguity [20]. This method conveniently combines the use of spatial and spectral (or intensity) information into one step. Joint methods using SVM fused output,

or SVM results combined with other image processing steps exist [21–23]. Other spatial techniques include clustering imagery previously exploited by other methods [15], morphological segmentation [24], including extracting the spatial information as a pre-processing step using watershed transformations [25].

The SIDE filter was briefly evaluated on real hyperspectral data in the context of the trained classifiers discussed in sec. 2.1 [26]. However, evaluating performance of detection techniques on real data can be an approximate process when exact ground truth is not available.

1.2 Summary

In Chapter 2 the concepts of detection and classification are discussed. Several existing techniques from these fields are explained, including some from the statistical learning literature [27, 28], and the clutter matched filter (CMF) [29], which is from the antenna literature (adapted here for the hyperspectral problem). In the case of vector-valued pixels in an image, the statistical learning techniques present a valuable tool when training data is available. In many instances, however, training data is not available. The clutter matched filter makes use of the foreknowledge gained in laboratory measurements of chemical spectral signatures, so it does not require training data to perform detection. Chapter 3 introduces the hyperspectral imaging paradigm and discusses the observation of gaseous plumes in long-wave infrared (LWIR) hyperspectral data. A physical model of plume observation is built from radiometric principles. The detection of these plumes is the main goal in the experiments of this work. The DIRSIG Model is also discussed in Chapter 3. DIRSIG was used to generate the data needed to quantify the performance of the techniques and tools presented in Chapter 2. Continuing to Chapter 4, the results of the experiments are presented. The data used for developing these experiments are examined as well as the results generated from several tests on the data. The new SIDE and bolt-on SIDE (B-SIDE) filters are quantitatively compared to existing techniques using the DIRSIG-generated synthetic images. Finally, Chapter 5 contains both the conclusions along with a discussion on possible future work in this area. Parameters of the SIDE filter are discussed,

as well as ways to combat the difficult problem of assigning realistic prior probabilities to neighborhoods of pixels. The contributions made in this thesis are stated in this final chapter. These include a full collection of results from experiments done using the SIDE and B-SIDE filters across many variables relating to physical conditions at the time images are collected. The B-SIDE filter is explored with exhaustive results for the first time in comparison to the CMF as well as statistical machine learning techniques. These show that the B-SIDE filter provides a valuable enhancement for detection using the spatial information readily available in images.

Chapter 2

Detection and Classification

The statistical classification of data has long been a problem that has received focused attention. The beginning of this work as it is done today is often attributed to Fisher's discrimination of plant species using a statistical method still in use today [30]. The discriminant described by Fisher in that early work will be discussed along with several others that have found importance in the field of classification. These techniques are presented here to provide some background into the classification problem and its various modes, as well as establish a baseline of existing algorithmic performance with which to compare the SIDE and B-SIDE filters, which are discussed later in this chapter. The following classification techniques are briefly treated in this section: Fisher's discriminant, the least squares linear classifier (LC), linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA).

The clutter matched filter (CMF), a common technique for detecting additive signals in Gaussian background, is briefly investigated. The CMF can be shown to be the optimal detector in data that fits a signal plus additive Gaussian background model. It has been applied with great success to vector-valued hyperspectral imaging where the data can fit that model. The SIDE filters, which are newly introduced, are fully developed and analyzed in sec. 2.3. These filters differ from the others introduced in this chapter because they incorporate the spatial information available in images into the detection problem. Two different approaches to this filter are discussed in this chapter, which are called the SIDE filter and the B-SIDE filter. The SIDE filter takes a matched filter style approach, where vector observations are assumed, whereas the B-SIDE filter takes a scalar input and takes a post-processing approach to detection enhancement. The efficient implementation of these two filters in MATLAB is briefly discussed for each.

2.1 Supervised Learning Techniques

Statistical learning for pattern recognition has been a popular technique for many years. The contributions made to this field date back to the "Father of Statistics," Sir Ronald A. Fisher, and his work in biology and genetics [30]. One particular group of techniques that have been developed through these years of work are the supervised learning techniques. These techniques can be trained using data that is known to belong to a certain class. Following the training step, the trained classifier (whether it be a vector of coefficients or a discriminating function) can be used to classify data that has not yet been encountered, but is assumed to belong to one of the classes for which the classifier was trained. The techniques described here exploit one or more the following concepts: Gaussianity of classes, minimization of an error metric, and maximization of a separation metric. The developments presented in this section draw heavily from the texts of Duda and Hart [28], as well as Hastie, et al. [27].

2.1.1 Fisher's Discriminant

The discriminant function described by Fisher [30] is the oldest classification technique used in this study. The goal of this discriminant is to project multi-class data ($\mathbf{x} \in \mathcal{X}_i, i \in \{0, 1, \ldots, k\}$) with high dimensionality to a low (perhaps one) dimensionality, while maintaining the separation of the classes. This discriminant is developed here in the simplest two-class case, where it is desired to project the data $\mathbf{x}$ onto a single dimension (by some $y = \mathbf{w}^T\mathbf{x}$) in such a way that maximizes the separation of the two classes. The set of projected data y could then be separated into two groups $\mathcal{Y}_1$ and $\mathcal{Y}_2$ that should correspond very closely to the original classes $\mathcal{X}_1$ and $\mathcal{X}_2$. Maximizing the separation of these two classes in the lower dimension can be done in a meaningful way by maximizing the difference between class means relative to some description of the variance of the classes. This is usually expressed as:

$$J(\mathbf{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}, \tag{2.1}$$

where the estimates of the class means (after being discriminated) are $\tilde{m}_1$ and $\tilde{m}_2$, while $\tilde{s}_1^2$

and $\tilde{s}_2^2$ denote class scatter. Scatter is an easily computed value that is related to variance, and so can be used here as a good metric with which to relate the difference of the means. The scatter is defined as

$$\tilde{s}_i^2 = \sum_{y \in \mathcal{X}_i} (y - \tilde{m}_i)^2. \tag{2.2}$$

The cost function (2.1) provides a meaningful statement of what is wished to be maximized. However, in order to train a classifying or separating vector $\mathbf{w}$, the function must be maximized with respect to $\mathbf{w}$. First we define the mean for class $\mathcal{X}_i$ as

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in \mathcal{X}_i} \mathbf{x}, \tag{2.3}$$

where n_i is the number of observations in class $\mathcal{X}_i$. Then, we can write the estimated class mean $\tilde{m}_i$ as

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in \mathcal{Y}_i} y = \frac{1}{n_i} \sum_{\mathbf{x} \in \mathcal{X}_i} \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{m}_i. \tag{2.4}$$

The numerator of the cost function can then be expressed in terms of $\mathbf{w}$ by first rewriting the scatter as

$$\begin{aligned} \tilde{s}_i^2 &= \sum_{\mathbf{x} \in \mathcal{X}_i} (\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \mathbf{m}_i)^2 \\ &= \sum_{\mathbf{x} \in \mathcal{X}_i} \mathbf{w}^T (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \mathbf{w} \\ &= \mathbf{w}^T S_i \mathbf{w}, \end{aligned} \tag{2.5}$$

and defining

$$S_i = \sum_{\mathbf{x} \in \mathcal{X}_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T. \tag{2.6}$$

Then, that definition for S_i is used in defining the within-class scatter matrix S_W:

$$S_W = S_1 + S_2. \tag{2.7}$$

The within-class scatter matrix is equal to a scalar multiple of the covariance matrix for

the entire data set, i.e., $S_W = (n_1 + n_2 - 1)K$. This allows the denominator of (2.1) to be expressed as

$$\tilde{s}_1^2 + \tilde{s}_2^2 = \mathbf{w}^T S_W \mathbf{w}. \tag{2.8}$$

The numerator of the cost function must also be expressed in terms of $\mathbf{w}$. This is done using (2.4) in the same way as in the denominator. The difference of means is written as

$$\begin{aligned} (\tilde{m}_1 - \tilde{m}_2)^2 &= (\mathbf{w}^T\mathbf{m}_1 - \mathbf{w}^T\mathbf{m}_2)^2 \\ &= \mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w} \\ &= \mathbf{w}^T S_B \mathbf{w}, \end{aligned} \tag{2.9}$$

where S_B is called the between-class scatter matrix and is defined

$$S_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T. \tag{2.10}$$

These expressions for the numerator and denominator of (2.1) can be combined as

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}. \tag{2.11}$$

This can be maximized by taking a derivative of (2.11):

$$\frac{\partial}{\partial \mathbf{w}} \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} = \frac{(\mathbf{w}^T S_W \mathbf{w})(2S_B\mathbf{w}) - (\mathbf{w}^T S_B \mathbf{w})(2S_W\mathbf{w})}{(\mathbf{w}^T S_W \mathbf{w})^2} = 0. \tag{2.12}$$

Setting this derivative equal to zero allows some of the quadratic scalar terms to be removed, so that

$$\frac{2S_B\mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} - \frac{(\mathbf{w}^T S_B \mathbf{w})(2S_W\mathbf{w})}{(\mathbf{w}^T S_W \mathbf{w})^2} = 0 \tag{2.13}$$

$$\frac{2S_B\mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} = \frac{(\mathbf{w}^T S_B \mathbf{w})}{(\mathbf{w}^T S_W \mathbf{w})^2}(2S_W\mathbf{w}). \tag{2.14}$$

Finally, this can be written as

$$S_B\mathbf{w} = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} S_W \mathbf{w}. \tag{2.15}$$

The resultant expression shows that the solution must take the form

$$S_B\mathbf{w} = \lambda S_W \mathbf{w}, \tag{2.16}$$

where $\lambda = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}$. This generalized eigenvalue problem is ameliorated by assuming that S_W is invertible, so that

$$S_W^{-1} S_B \mathbf{w} = \lambda \mathbf{w}, \tag{2.17}$$

which is now a standard eigenvalue problem. The problem is simplified even further by noticing that $S_B\mathbf{w}$ can be written

$$(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{w} = (\mathbf{m}_1 - \mathbf{m}_2)(\tilde{m}_1 - \tilde{m}_2), \tag{2.18}$$

which is always a scalar multiple (and so in the direction) of the difference in the class means $(\mathbf{m}_1 - \mathbf{m}_2)$ [28]. Since the scale factor (λ) on $\mathbf{w}$ is of no consequence, the maximizing solution can be written as

$$\mathbf{w} = S_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2). \tag{2.19}$$

2.1.2 Linear Regression

The least squares linear classifier is perhaps the simplest of all when approached from the linear algebra perspective. The data is assumed to be two-class, and to fit a linear model,

$$y = \rho_0 + \sum_{i=1}^{m} \rho_i x_i, \tag{2.20}$$

where the data are x_i and the coefficients of a classifying linear function, ρ_i, aid in producing the classification result y. This is more conveniently written as

$$y = \mathbf{x}^T \boldsymbol{\rho}, \tag{2.21}$$

where a vector of the x_i has been augmented with a constant term 1 to capture ρ_0, making $\mathbf{x}, \boldsymbol{\rho} \in \mathbb{R}^{m+1}$. This allows the linear model to be expressed in concise vector notation. A minimum mean square error (MMSE) approach is used to estimate the coefficients of the classifying vector ($\hat{\boldsymbol{\rho}}$) and consequently obtain estimates for the class labels ($\hat{\mathbf{y}}$) for each data point. Now suppose there are n observations of $\mathbf{x}$. Let $\mathbf{X} \in \mathbb{R}^{n \times m+1}$ be the observations stacked in rows. Let $\mathbf{y}$ represent the known class labels for these training data. The MMSE approach can then be expressed as

$$\hat{\boldsymbol{\rho}} = \min_{\boldsymbol{\rho}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\rho}\|^2 . \tag{2.22}$$

The usual answer is a pseudoinverse of $\mathbf{X}$, if $\mathbf{X}$ is not square and invertible:

$$\hat{\boldsymbol{\rho}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}\mathbf{y}. \tag{2.23}$$

This final statement shows that training data contained in $\mathbf{y}$ and $\mathbf{X}$ can be used to calculate an estimate of $\boldsymbol{\rho}$ in the linear model.

2.1.3 Linear Discriminant Analysis (LDA)

Linear discriminant analysis is yet another supervised learning technique that can be thought of as a projection onto a single dimension from any higher dimension. This allows the final step of thresholding and classification to be a simple one. The development of LDA begins with an assumption about the data that will become very important later on in this chapter: Gaussianity. The data, once again $\mathbf{x}$, is assumed to be a Gaussian random vector where $\mathbf{x} \in \mathbb{R}^m$. Let there be two classes among the data, so that each observation

$\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}, K)$, $i \in \{0, 1\}$. The probability density function of each class can be written as

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{m/2}\sqrt{|K_i|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T K_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right). \tag{2.24}$$

Also, let observations from each class occur with some prior probability, π_i. The linear discriminant functions then arise from a log-likelihood ratio (LLR) of class posterior likelihood functions when the class covariance matrices K_i are assumed to be equal, $K = K_i$, $\forall\, i \in \{0, 1\}$. This ratio,

$$\log \frac{P(i|\mathbf{x})}{P(j|\mathbf{x})}, \tag{2.25}$$

can be written using Bayes' Rule as

$$\log \frac{P(i|\mathbf{x})}{P(j|\mathbf{x})} = \log \frac{f_i(\mathbf{x})}{f_j(\mathbf{x})} + \log \frac{\pi_i}{\pi_j}. \tag{2.26}$$

The expansion of the LLR of the density functions reveals a structure that can be separated into two different functions:

$$\log \frac{P(i|\mathbf{x})}{P(j|\mathbf{x})} = -\frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)^T K^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \mathbf{x}^T K^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \log \frac{\pi_i}{\pi_j}, \tag{2.27}$$

$$\log \frac{P(i|\mathbf{x})}{P(j|\mathbf{x})} = \left(-\frac{1}{2}\boldsymbol{\mu}_i^T K^{-1}\boldsymbol{\mu}_i + \mathbf{x}^T K^{-1}\boldsymbol{\mu}_i + \log \pi_i\right) - \left(-\frac{1}{2}\boldsymbol{\mu}_j^T K^{-1}\boldsymbol{\mu}_j + \mathbf{x}^T K^{-1}\boldsymbol{\mu}_j + \log \pi_j\right). \tag{2.28}$$

At this point the LLR is usually split into the two linear discriminant functions, which are written as

$$g_i(\mathbf{x}) = -\frac{1}{2}\boldsymbol{\mu}_i^T K^{-1}\boldsymbol{\mu}_i + \mathbf{x}^T K^{-1}\boldsymbol{\mu}_i + \log \pi_i. \tag{2.29}$$

The actual application of these functions in decision-making requires that some comparison be made between the two functions for a given observation. Obviously, these two can be differenced, as the original LLR (2.26) demands:

$$\log \frac{P(i|\mathbf{x})}{P(j|\mathbf{x})} = g_i(\mathbf{x}) - g_j(\mathbf{x}). \tag{2.30}$$

In this configuration, the threshold for decision must be 0. When stated this way, (2.30), a change in the class priors will not change the threshold for decision. A drawback to the LDA technique is that it is a simple linear function, so decision boundaries can only be straight lines in the original space. This can be a limitation in performance if the data exhibits some more interesting behaviors, such as curved boundaries.

2.1.4 Quadratic Discriminant Analysis (QDA)

Quadratic discriminant functions arise when the assumption that the classes share a covariance matrix, K, is thrown out. Let each class have a distinct covariance matrix K_k. The expansion and simplification of the LLR from (2.26) takes on a slightly different form. This becomes:

$$\log \frac{P(i|\mathbf{x})}{P(j|\mathbf{x})} = \log \left[\sqrt{\frac{(2\pi)^m |K_j|}{(2\pi)^m |K_i|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T K_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^T K_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j) \right) \right] + \log \frac{\pi_i}{\pi_j}. \tag{2.31}$$

The simplification of the terms that depend on $\mathbf{x}$ begins with

$$\log \frac{f_j(\mathbf{x})}{f_j(\mathbf{x})} = \log \left[\sqrt{\frac{|K_j|}{|K_i|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T K_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^T K_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j) \right) \right], \tag{2.32}$$

and application of the log function throughout the expression shows the two-functional nature of this ratio:

$$\begin{aligned}\log\frac{f_j(\mathbf{x})}{f_j(\mathbf{x})} &= \frac{1}{2}\log|K_j| - \frac{1}{2}\log|K_i| - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T K_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^T K_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j) \\ &= -\frac{1}{2}\log|K_i| - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T K_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \frac{1}{2}\log|K_j| + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^T K_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j) \\ &= \tilde{g}_i(\mathbf{x}) - \tilde{g}_j(\mathbf{x}), \end{aligned} \tag{2.33}$$

where $\tilde{g}_i(\mathbf{x})$ is part of the quadratic discriminant function (the priors are missing). This can be used to split the LLR (2.31) into two functions, such as in the LDA case. The priors being reincluded,

$$g_i(\mathbf{x}) = -\frac{1}{2}\log|K_i| - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T K_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \log\pi_i. \tag{2.34}$$

These functions are the quadratic discriminant functions. They are called the "quadratic" functions because they are quadratic in the observation term $\mathbf{x}$. A similar differencing and thresholding approach to the one for LDA can be used when applying the quadratic discriminant functions to a classification or decision problem. The QDA approach to classification can fit to curved boundaries in the original data space. This can be a very desirable trait when fitting to data that are not separated simply by a difference in mean.

2.2 Clutter Matched Filter

The CMF is a recognized method and provides a standard by which to compare other detection techniques in the vector-valued observation paradigm. Consider the vector-valued signal model

$$\mathbf{r}_t = \mathbf{z}_t + \epsilon_t \mathbf{b}, \tag{2.35}$$

where the pixels in an image are indexed by t and the background image information is $\mathbf{z}$. The vectors $\mathbf{r}_t, \mathbf{z}_t, \mathbf{b}$ are $m \times 1$ column vectors. We assume $\mathbf{z}_t \sim \mathcal{N}(\boldsymbol{\mu}_z, K)$. A signal of interest is denoted by $\mathbf{b}$ and a signal "strength" is denoted by ϵ_t. In this situation it may be desired to detect the presence of $\mathbf{b}$ among the background information.

The clutter matched filter is an example of an optimal detector for an additive signal on a gaussian background in vector-valued data. The clutter matched filter is commonly written as

$$CMF_t = \frac{\mathbf{b}^T K^{-1} \mathbf{r}_t}{\sqrt{\mathbf{b}^T K^{-1} \mathbf{b}}}. \tag{2.36}$$

There are several ways to arrive at this expression [31]. One method is to maximize a signal-to-cluter ratio much like the one maximized for Fisher's discriminant in (2.1). This method uses a similar approach to the one taken in sec. 2.1.1, except to find the maximum of the signal-to-clutter ratio, Lagrange multipliers are used and the minimum of a constrained optimization cost function is found. Another common method is to assume the background data is Gaussian — as has been done for (2.35) — and perform a hypothesis test on the presence of the signal $\mathbf{b}$ in the observation. This is the approach that will be taken here.

Let the observations have this probability density function:

$$f(\mathbf{r}) = \frac{1}{(2\pi)^{m/2}\sqrt{|K|}} \exp\left(-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^T K^{-1}(\mathbf{r}-\boldsymbol{\mu})\right), \tag{2.37}$$

where $\boldsymbol{\mu}$ is the mean and K is the covariance. Propose a hypothesis test on the presence of signal in the observation. This really tests for a shift in the mean, as shown in the hypotheses below. Let

$$\begin{aligned} H_0 &: \boldsymbol{\mu} = \boldsymbol{\mu}_z \\ H_1 &: \boldsymbol{\mu} = \boldsymbol{\mu}_z + \epsilon\mathbf{b}, \end{aligned} \tag{2.38}$$

and form the likelihood ratio for this test. This likelihood ratio can be written as

$$\mathcal{L}(\mathbf{r}) = \frac{\exp\left[-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}-\epsilon\mathbf{b})^T K^{-1}(\mathbf{r}-\boldsymbol{\mu}-\epsilon\mathbf{b})\right]}{\exp[-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^T K^{-1}(\mathbf{r}-\boldsymbol{\mu})]}. \tag{2.39}$$

The negative log of the LR can be expressed

$$\begin{aligned}
-\log \mathcal{L}(\mathbf{r}) &= \frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}-\epsilon\mathbf{b})^T K^{-1}(\mathbf{r}-\boldsymbol{\mu}-\epsilon\mathbf{b}) - \frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^T K^{-1}(\mathbf{r}-\boldsymbol{\mu}) \\
&= \frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^T K^{-1}(\mathbf{r}-\boldsymbol{\mu}-\epsilon\mathbf{b}) - \frac{1}{2}\epsilon\mathbf{b}^T K^{-1}(\mathbf{r}-\boldsymbol{\mu}-\epsilon\mathbf{b}) \\
&\quad - \frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^T K^{-1}(\mathbf{r}-\boldsymbol{\mu}) \\
&= \frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^T K^{-1}(\mathbf{r}-\boldsymbol{\mu}) - \frac{1}{2}\epsilon(\mathbf{r}-\boldsymbol{\mu})^T K^{-1}\mathbf{b} - \frac{1}{2}\epsilon\mathbf{b}^T K^{-1}(\mathbf{r}-\boldsymbol{\mu}) \\
&\quad - \frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^T K^{-1}(\mathbf{r}-\boldsymbol{\mu}) + \frac{1}{2}\epsilon\mathbf{b}^T K^{-1}\mathbf{b} \\
&= -\epsilon\mathbf{b}^T K^{-1}(\mathbf{r}-\boldsymbol{\mu}) + \frac{1}{2}\epsilon\mathbf{b}^T K^{-1}\mathbf{b} \\
&= -\epsilon\mathbf{b}^T K^{-1}\mathbf{r} + \epsilon\mathbf{b}^T K^{-1}\boldsymbol{\mu} + \frac{1}{2}\epsilon\mathbf{b}^T K^{-1}\mathbf{b}, \qquad (2.40)
\end{aligned}$$

which shows that this LLR can be thought of as a linear function in $\mathbf{r}$. The final form of the LLR shows a single multiply on $\mathbf{r}$ and two constant additive terms. That is, if $\mathbf{b}^T K^{-1}\mathbf{r}$ is large, then the likelihood of plume present in that observation is large. The additive terms can be ignored because they only shift the threshold, and the scaling by ϵ is immaterial as well. This derivation reveals the arbitrary nature of the scaling factor in (2.36). That scaling term is a result of the signal-to-cluter maximization approach to deriving the CMF. In forming the Lagrangian that maximizes the signal-to-cluter ratio, $\mathbf{b}^T K^{-1}\mathbf{b}$ is held constant as the constraint, and the value chosen decides that scaling factor.

2.3 SIDE Filter

The purpose of the SIDE filter is to exploit the spatial information discussed in Chapter 1. This information is available, given the assumptions made in that same section. A probabilistic approach is taken to extracting and utilizing that information. The realization of the filter is developed in the case of vector-valued observations, and then the case of scalar-valued observations is treated.

2.3.1 Vector-Valued SIDE

Let the signal model (2.35) be augmented with a new term indicating the presence of signal, so that the model is now

$$\mathbf{r}_t = \mathbf{z}_t + \epsilon_t s_t \mathbf{b}, \tag{2.41}$$

and let $s_t \in \{0, 1\}$, where a 1 indicates the presence of signal $\mathbf{b}$ in pixel t and 0 indicates an absence of $\mathbf{b}$ in pixel t. The s_t variable allows the signal model to describe a neighborhood structure.

Assuming the data have sufficient resolution (see Chapter 1), in small neighborhoods it is more probable that the signal $\mathbf{b}$ will be either uniformly present or absent or if the neighborhood includes a boundary, the boundaries are likely to be smooth. Figure 2.1 shows three different neighborhoods with light and dark pixels indicating the absence and presence of signal (plume), respectively. Figure 2.1(a) shows a region without a plume, fig. 2.1(b) shows a region in the interior of a plume, fig. 2.1(c) presents a region with a plume edge, while fig. 2.1(d) shows a random configuration that is unlikely based on physical processes, assuming the plume is large compared to the pixels. Based on plume behavior, figs. 2.1(a) and 2.1(b) are the most likely regions, fig. 2.1(c) a boundary region is less likely, but more likely than the random locations in fig. 2.1(d). These basic relative probabilities (or likelihoods) model the spatial continuity of the plume, the property we are exploiting.

Let t_0 be a pixel of interest, and N_{t_0} denote a pixel neighborhood surrounding t_0, not

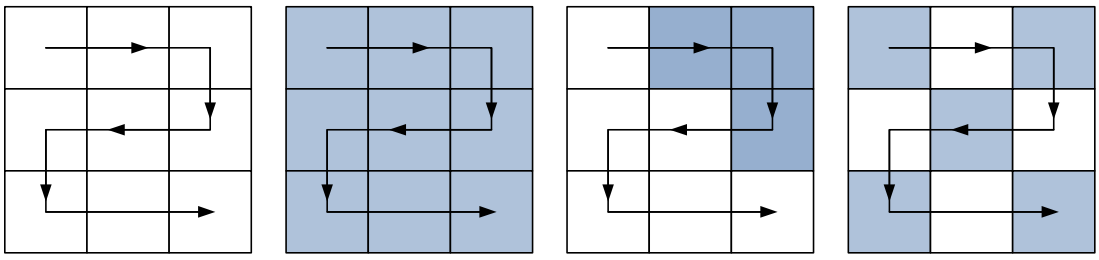
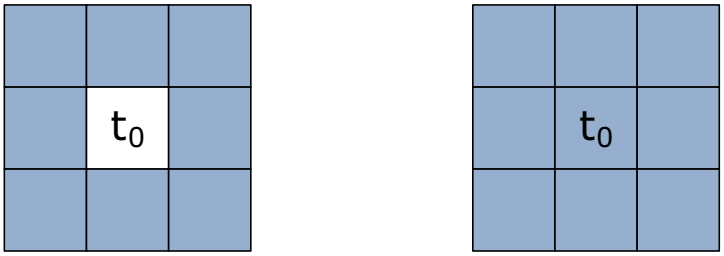


Fig. 2.1: Several neighborhood configurations and their corresponding raster scan value for $\Sigma\Delta(\mathbf{s})$ are shown (a) $\Sigma\Delta(\mathbf{s}) = 0$, (b) $\Sigma\Delta(\mathbf{s}) = 0$, (c) $\Sigma\Delta(\mathbf{s}) = 2$, (d) $\Sigma\Delta(\mathbf{s}) = 8$.



(a) (b)

Fig. 2.2: Neighborhoods demonstrating the N notation: (a) N, and (b) N^*.

including t_0 as shown in fig. 2.2(a). Let $N^*_{t_0}$ denote the set of pixels in the neighborhood of t_0 including t_0, the neighborhood shown in fig. 2.2(b).

The neighborhood shape and size are user-defined and can be chosen to serve a specific purpose. For example, fig. 2.3 shows neighborhoods that emphasize radial relationships (fig. 2.3(a)) and horizontal or vertical extent (fig. 2.3(c)). This work primarily uses a 3×3 neighborhood with the pixel of interest at the center of the neighborhood (figs. 2.1 and 2.2), but several results from a test using the 13-pixel neighborhood in fig. 2.3(a) are also presented in Chapter 4.

This allows the development of a set of indicator functions s_t for $t \in N^*$ for any small neighborhood signal configuration around pixel t_0. This set is called $\mathbf{S}$ and can be written $\mathbf{S}_{t_0} = \{s_t, t \in N^*_{t_0}\} \in \{0,1\}^{|N^*_{t_0}|}$. It is important to distinguish a specific neighborhood configuration of signal presence or absence from its set of indicator functions as $\mathbf{s} = \{s_t, t \in N^*_{t_0}\}$ (not dependent on t_0). There exists a prior probability for each of these neighborhood signal configurations based on the assumptions in Chapter 1. Let $P_{\mathbf{S}}(\mathbf{s})$ denote those priors. They can be calculated or assigned depending on the application. Probabilities could be modeled or computed for each neighborhood configuration or state using a physical or mathematical description. It is also possible to assign configuration probabilities to neighborhoods with certain orientations or particular shapes.

In this work is proposed and tested a simple, easily computed probability assignment that is consistent with the physical process. With the original assumptions — signal presence tends to be contiguous and data have sufficient resolution — the priors can be modeled

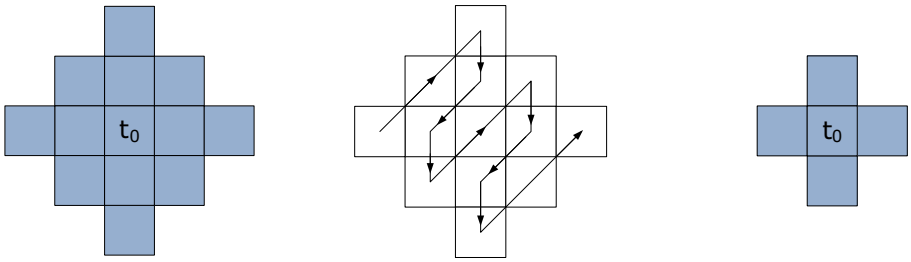


Fig. 2.3: Possible neighborhood configurations. (a) An example of a larger neighborhood that might emphasize radial relationships, along with (b) an example of a raster scan for this neighborhood shape, and (c) an example of a neighborhood that might be chosen to emphasize direct contact between pixels.

by performing a raster scan of a given neighborhood and counting the changes in the indicator values (fig. 2.1). This measure, called $\Sigma\Delta(\mathbf{s})$, can be used in a prior model where

$$P_{\mathbf{S}}(\mathbf{s}) \propto e^{-\alpha\Sigma\Delta(\mathbf{s})}. \tag{2.42}$$

$P_{\mathbf{S}}(\mathbf{s})$ is a generally decreasing function of $\Sigma\Delta(\mathbf{s})$. Figure 2.4 shows a possible assignment for $P_{\mathbf{S}}(\mathbf{s})$. The α parameter determines the relative emphasis between small $\Sigma\Delta(\mathbf{s})$ values and large $\Sigma\Delta(\mathbf{s})$ values, representing the relative probability of these spatial distributions. Smaller $\Sigma\Delta(\mathbf{s})$ values represent neighborhoods with greater contiguity and smoother boundaries. Larger $\Sigma\Delta(\mathbf{s})$ values result from more random spatial distributions where the high spatial frequency of signal presence causes this measure of "changes" in a neighborhood to increase. A small α value (e.g., $\alpha \in [0, 1]$) represents less of a difference in probability between less contiguous and more contiguous neighborhoods, while a large α value emphasizes the probability difference between contiguous and less contiguous neighborhoods.

This raster scan method is rotationally variant for some neighborhood configurations. To remove that variance, the scan could be done multiple times, each time rotating the neighborhood by 90°. These counts could then be averaged to obtain a rotationally invariant measure. Other methods to assign prior probabilities could also be used, for example, they could be estimated based on existing data with similar features.

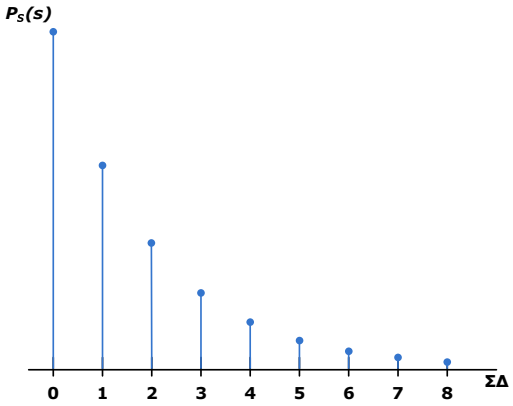


Fig. 2.4: A generally decreasing example for $P_{\mathbf{S}}(\mathbf{s})$ as a function of $\Sigma\Delta(\mathbf{s})$.

Hypothesis Testing

Form the hypothesis test on the presence of plume in the pixel of interest, that is, pixel t_0 as

$$\begin{aligned} H_0 &: s_{t_0} = 0 \\ H_1 &: s_{t_0} = 1. \end{aligned} \tag{2.43}$$

By the Law of Total Probability [32], the densities for each hypothesis case can be written

$$p_h(\mathbf{y}_{t_0}) = \sum_{\mathbf{s}:s_{t_0}=h} p_{\mathbf{Y}}(\mathbf{y}_{t_0}|\mathbf{s}) P_{\mathbf{S}}(\mathbf{s}), \tag{2.44}$$

where h is the current hypothesis, $h \in \{0, 1\}$, and $P_{\mathbf{S}}(\mathbf{s})$ is the prior neighborhood probability from (2.42). The joint vector of observations in $N^*_{t_0}$ is $\mathbf{y}_{t_0}$, where $\mathbf{y}_{t_0} = \{\mathbf{r}_t : t \in N^*_{t_0}\}$. The signal strength term is assumed for simplicity to be constant in the current neighborhood, that is, $\epsilon_t = \epsilon, t \in N^*_{t_0}$. Assuming that the noise components in the pixels in the neighborhood are independent of each other probabilistically, and using (2.41), the conditional distributions from the statement above are written as

$$p_{\mathbf{Y}}(\mathbf{y}_{t_0}|\mathbf{s}) = \prod_{t \in N^*_{t_0}} \mathcal{N}(\mathbf{r}_t; \epsilon s_t \mathbf{b}, K), \tag{2.45}$$

where $\mathcal{N}(\mathbf{r}_t; \epsilon s_t \mathbf{b}, K)$ is the normal distribution for random vector $\mathbf{r}_t$ with mean $\epsilon s_t \mathbf{b}$ and covariance K. The sample covariance matrix associated with the image data is used for K.

The likelihood ratio (LR) ϕ_{t_0} for the hypothesis test (2.43) is then calculated as

$$\phi_{t_0} = \frac{p_1(\mathbf{y}_{t_0})}{p_0(\mathbf{y}_{t_0})} = \frac{\sum_{\mathbf{s}:s_{t_0}=1} p_{\mathbf{Y}}(\mathbf{y}_{t_0}|\mathbf{s}) P_{\mathbf{S}}(\mathbf{s})}{\sum_{\mathbf{s}:s_{t_0}=0} p_{\mathbf{Y}}(\mathbf{y}_{t_0}|\mathbf{s}) P_{\mathbf{S}}(\mathbf{s})}, \tag{2.46}$$

which can be written as

$$\phi_{t_0} = \frac{\sum_{\mathbf{s}:s_{t_0}=1} \prod_{t\in N^*_{t_0}} \exp\left[-\frac{1}{2}\left(\mathbf{r}_t - \epsilon s_t \mathbf{b}\right)^T K^{-1}\left(\mathbf{r}_t - \epsilon s_t \mathbf{b}\right)\right] P_{\mathbf{S}}(\mathbf{s})}{\sum_{\mathbf{s}:s_{t_0}=0} \prod_{t\in N^*_{t_0}} \exp\left[-\frac{1}{2}\left(\mathbf{r}_t - \epsilon s_t \mathbf{b}\right)^T K^{-1}\left(\mathbf{r}_t - \epsilon s_t \mathbf{b}\right)\right] P_{\mathbf{S}}(\mathbf{s})} \tag{2.47}$$

$$= \frac{\exp\left[-\frac{1}{2}\sum_{t\in N^*_{t_0}} \mathbf{r}_t^T K^{-1} \mathbf{r}_t\right] \sum_{\mathbf{s}:s_{t_0}=1} \exp\left[\epsilon \sum_{t\in N^*_{t_0}} \mathbf{r}_t^T K^{-1}\mathbf{b} s_t - \frac{1}{2}\epsilon^2 \sum_{t\in N^*_{t_0}} \mathbf{b}^T K^{-1}\mathbf{b} s_t\right] P_{\mathbf{S}}(\mathbf{s})}{\exp\left[-\frac{1}{2}\sum_{t\in N^*_{t_0}} \mathbf{r}_t^T K^{-1} \mathbf{r}_t\right] \sum_{\mathbf{s}:s_{t_0}=0} \exp\left[\epsilon \sum_{t\in N^*_{t_0}} \mathbf{r}_t^T K^{-1}\mathbf{b} s_t - \frac{1}{2}\epsilon^2 \sum_{t\in N^*_{t_0}} \mathbf{b}^T K^{-1}\mathbf{b} s_t\right] P_{\mathbf{S}}(\mathbf{s})} \tag{2.48}$$

$$\phi_{t_0} = \frac{\sum_{\mathbf{s}:s_{t_0}=1} \exp\left[\left(\epsilon \sum_{t\in N^*_{t_0}} m_t s_t\right) - \left(\beta^2\epsilon^2/2 \sum_{t\in N^*_{t_0}} s_t\right)\right] P_{\mathbf{S}}(\mathbf{s})}{\sum_{\mathbf{s}:s_{t_0}=0} \exp\left[\left(\epsilon \sum_{t\in N^*_{t_0}} m_t s_t\right) - \left(\beta^2\epsilon^2/2 \sum_{t\in N^*_{t_0}} s_t\right)\right] P_{\mathbf{S}}(\mathbf{s})} \tag{2.49}$$

$$\phi_{t_0} = \frac{\sum_{\mathbf{s}:s_{t_0}=1} \exp\left[\epsilon \sum_{t\in N^*_{t_0}} s_t \left(m_t - \frac{\beta^2\epsilon}{2}\right)\right] P_{\mathbf{S}}(\mathbf{s})}{\sum_{\mathbf{s}:s_{t_0}=0} \exp\left[\epsilon \sum_{t\in N^*_{t_0}} s_t \left(m_t - \frac{\beta^2\epsilon}{2}\right)\right] P_{\mathbf{S}}(\mathbf{s})}, \tag{2.50}$$

where $m_t = \mathbf{b}^T K^{-1}\mathbf{r}_t$ and $\beta^2 = \mathbf{b}^T K^{-1}\mathbf{b}$. The quantity m_t is similar to the clutter matched filter (CMF) [7, 29] with respect to $\mathbf{b}$ for pixel t. The CMF provides a useful baseline for evaluating the performance of SIDE.

This test reduces to the usual LR test in the case that there is only one pixel in the neighborhood, or in the case that the prior probability distribution $P_{\mathbf{S}}(\mathbf{s})$ is constant over

all **s**.

Computations

Although the expression for the LR in (2.50) appears simple and is a convenient way to implement the SIDE filter in MATLAB, there is a degree of redundancy in computation that can be reduced with a little further inspection. The summations over **s** are expanded in the numerator and denominator:

$$\phi_{t_0} = \frac{d_0 p_{1,0} + d_1 p_{1,1} + \cdots + d_n p_{1,n}}{a_0 p_{0,0} + a_1 p_{0,1} + \cdots + a_n p_{0,n}}, \tag{2.51}$$

where the indices for d_i and a_i correspond to neighborhood configurations $\mathbf{s}_{1,i}$ and $\mathbf{s}_{0,i}$ (where the 0 and 1 correspond to the hypothesis value from (2.43)) that are exactly the same except for the pixel t_0. That is, for d_i, $s_{t_0} = 1$ and for a_i, $s_{t_0} = 0$, but the configurations are otherwise the same. The prior probabilities are represented by $p_{h,i}$, where h is the hypothesis value from (2.43). The following are the meanings for each symbol:

$$d = \exp\left[\epsilon \sum_{t \in N^*_{t_0}} s_t \left(m_t - \frac{\epsilon\beta^2}{2}\right)\right], \tag{2.52}$$

$$a = \exp\left[\epsilon \sum_{t \in N_{t_0}} s_t \left(m_t - \frac{\epsilon\beta^2}{2}\right)\right], \tag{2.53}$$

$$p_{h,i} = P_{\mathbf{S}}(\mathbf{s}_{h,i}). \tag{2.54}$$

The sum in a does not include pixel t_0 because the term corresponding to that pixel is 0. The simplification that can now be made is to expand d, which does include the term

corresponding to the center pixel:

$$\begin{aligned} d &= \exp\left[\epsilon\left(m_{t_0} - \frac{\epsilon\beta^2}{2}\right)\right] \exp\left[\epsilon \sum_{t\in N_{t_0}} s_t \left(m_t - \frac{\epsilon\beta^2}{2}\right)\right] \\ &= \exp\left[\epsilon\left(m_{t_0} - \frac{\epsilon\beta^2}{2}\right)\right] a \\ &= d_e a. \end{aligned} \tag{2.55}$$

This shows that the particular information d_e can be separated from the spatial information a and that this spatial information is present in the numerator and denominator of the LR. The LR can now be written in a more computationally efficient way:

$$\phi_{t_0} = \frac{d_e(a_0 p_{1,0} + a_1 p_{1,1} + \cdots + a_n p_{1,n})}{a_0 p_{0,0} + a_1 p_{0,1} + \cdots + a_n p_{0,n}} \tag{2.56}$$

$$= \exp\left[\epsilon\left(m_{t_0} - \frac{\epsilon\beta^2}{2}\right)\right] \frac{\displaystyle\sum_{\mathbf{s}:s_{t_0}=1} \exp\left[\epsilon \sum_{t\in N_{t_0}} s_t \left(m_t - \frac{\epsilon\beta^2}{2}\right)\right] P_{\mathbf{S}}(\mathbf{s})}{\displaystyle\sum_{\mathbf{s}:s_{t_0}=0} \exp\left[\epsilon \sum_{t\in N_{t_0}} s_t \left(m_t - \frac{\epsilon\beta^2}{2}\right)\right] P_{\mathbf{S}}(\mathbf{s})}. \tag{2.57}$$

This form allows the spatial information in a to be calculated just once. While the sum must be calculated for both the numerator and denominator of this LR because of the independent priors, the work can be simplified in this manner.

2.3.2 B-SIDE

The result of an initial exploitation step, such as an unscaled matched filter, denoted m_t, is the input for the LR test (2.50). For this application, the LR is a two-step process: a processing step that reduces the vector image to a scalar image (presence or absence of signal) then application of SIDE to exploit the spatial information. This allows the use of this same approach for scalar-valued image data in a way that resembles a nonlinear filter that enhances pixels that are similar over neighborhoods.

With this insight, SIDE can be viewed as a post-processing filter for image analysis,

which will be called the Bolt-on SIDE, or B-SIDE. Any other detectors can be used that provide a scalar image C with pixels c_t for hypothesis testing and after initial processing, the image C is passed to the B-SIDE processor to exploit the spatial information. This requires there be an input image, C, with two pixel classes representing the presence or absence of signal:

$$\begin{aligned} &\text{Signal absent}: \quad c_t \sim \mathcal{N}(\mu_0, \sigma_c^2) \\ &\text{Signal present}: \quad c_t \sim \mathcal{N}(\mu_1, \sigma_c^2). \end{aligned} \tag{2.58}$$

These two pixel classes can be placed in a signal model much like the vector case:

$$c_t = z_t + s_t b, \tag{2.59}$$

where $z_t \sim \mathcal{N}(\mu_0, \sigma_c^2)$ and b is a constant value such that $\mu_1 = \mu_0 + b$. The indicator s_t indicates the presence of this signal which determines the pixel class similar to the vector case. Using this same model, let $\phi_{t_0}^s$ denote the LR for the hypothesis test using this new signal model and write

$$\phi_{t_0}^s = \frac{\displaystyle\sum_{\mathbf{s}:s_{t_0}=1} \exp\left[\frac{1}{\sigma_c^2}\left(\sum_{t\in N_{t_0}^*} c_t\mu_{s_t} - \frac{1}{2}\sum_{t\in N_{t_0}^*}\mu_{s_t}^2\right)\right] P_{\mathbf{S}}(\mathbf{s})}{\displaystyle\sum_{\mathbf{s}:s_{t_0}=0} \exp\left[\frac{1}{\sigma_c^2}\left(\sum_{t\in N_{t_0}^*} c_t\mu_{s_t} - \frac{1}{2}\sum_{t\in N_{t_0}^*}\mu_{s_t}^2\right)\right] P_{\mathbf{S}}(\mathbf{s})}, \tag{2.60}$$

where μ_{s_t} is the mean of the class associated with s_t. This LR can be used as previously discussed for (2.50), but now using images pre-processed by a detection algorithm that classified the pixels according to the class description in (2.58). The same simplifications that motivated (2.49) to (2.50) can be used to simplify (2.60). The resultant expression,

$$\phi_{t_0}^s = \frac{\displaystyle\sum_{\mathbf{s}:s_{t_0}=1} \exp\left[\frac{\mu_1-\mu_0}{\sigma_c^2}\sum_{t\in N_{t_0}^*} s_t\left(c_t - \frac{\mu_1+\mu_0}{2}\right)\right] P_{\mathbf{S}}(\mathbf{s})}{\displaystyle\sum_{\mathbf{s}:s_{t_0}=0} \exp\left[\frac{\mu_1-\mu_0}{\sigma_c^2}\sum_{t\in N_{t_0}^*} s_t\left(c_t - \frac{\mu_1+\mu_0}{2}\right)\right] P_{\mathbf{S}}(\mathbf{s})}, \tag{2.61}$$

requires the substitution of μ_{s_t} with $\mu_0 + s_t b$, and a further step of substituting $\mu_1 - \mu_0$ for b to arrive at (2.61). This expression appears more complicated on the page, but can be a more efficient way to describe the B-SIDE filter in MATLAB. This efficiency can be achieved by using this expression to dictate pre-calculation of terms before the potentially long LR loop is computed.

A further simplification similar to the one accomplished in (2.57) can be done with the B-SIDE, where the final result is

$$\phi^s_{t_0} = \exp\left[\frac{\mu_1-\mu_0}{\sigma_c^2}\left(c_{t_0} - \frac{\mu_1+\mu_0}{2}\right)\right] \frac{\displaystyle\sum_{\mathbf{s}:s_{t_0}=1} \exp\left[\frac{\mu_1-\mu_0}{\sigma_c^2}\sum_{t\in N_{t_0}} s_t\left(c_t - \frac{\mu_1+\mu_0}{2}\right)\right] P_{\mathbf{S}}(\mathbf{s})}{\displaystyle\sum_{\mathbf{s}:s_{t_0}=0} \exp\left[\frac{\mu_1-\mu_0}{\sigma_c^2}\sum_{t\in N_{t_0}} s_t\left(c_t - \frac{\mu_1+\mu_0}{2}\right)\right] P_{\mathbf{S}}(\mathbf{s})}. \tag{2.62}$$

The resultant expression has the same benefits as the one for the SIDE filter in (2.57): the terms that correspond to spatial information for each pixel appear both in the numerator and denominator, and using this expression, it is straightforward to perform those computations just once.

Chapter 3

Hyperspectral Imaging

"Hyperspectral imagery is big. You might think space is big, but space is only three dimensional." *James Theiler, Bernard Foy, and Andrew Fraser*

Hyperspectral imaging is a relatively new concept, and is an abundant data source. So much information is captured at once that hardware has only recently been able to process it within the limits of human patience. This chapter will present a brief introduction to hyperspectral imaging, where some important hyperspectral sensors and their specifications are mentioned. Hyperspectral data is the application chosen for testing the SIDE filter and B-SIDE filter for this study. The model for gaseous plumes in hyperspectral data is developed in a wavelength-continous setting and then migrated to the discrete wavelengths that hyperspectral imagers operate in. This model is compared to the linear signal model in (2.35) that is assumed by the CMF and SIDE filter. This model is the basis for the development of those filters, so understanding it is crucial to applying them successfully. Finally, the DIRSIG Model, which was used to generate the data used in this study, is explained on a high level. The DIRSIG Model is a valuable tool which made quantification of the results in this study a realizable task.

3.1 Imaging Beyond Intensity

To demonstrate the utility of SIDE, it will be applied to the problem of identifying and characterizing weak chemical plumes in hyperspectral image data. The CMF will then be used as a baseline for comparison of performance.

Hyperspectral imaging sensors used in the remote sensing field generally collect spectral information from the long-wave infrared through ultraviolet regions. Hyperspectral imaging is not strictly defined, but it generally refers to data with dozens to hundreds of spectral

channels. As opposed to multi-spectral sensors, where spectral bands are not continuous and are selected to exploit particular information, hyperspectral sensors record continuous spectra described by the spectral resolution of the instrument.

Hyperspectral imaging has been applied to a number of different problems in remote sensing. One of these is the detection, identification, and characterization of chemical plumes in the scene. Successful gas plume detection in images uses techniques that match material library spectra to identify and characterize the gas using modified GLS approaches as discussed in sec. 1.1. These techniques have more in common with traditional chemical spectroscopy (e.g., matching library spectra) than traditional remote sensing (e.g., end-member extraction). The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [33,34] has been used in a number of these studies. AVIRIS is representative of a class of imagers that operate in one or more of the visible (VIS), near infrared (NIR), short-wave infrared (SWIR), or long-wave infreared (LWIR) spectral regions [14]. In the LWIR spectral division, materials are generally emissive (glow) and gases have unique spectral features that can be used for identification. In the shorter wavelengths, where plume emission is not a dominant factor in what is observed by the sensor, gas plumes have a predominantly absorptive effect that can be detected [35]. AVIRIS operates in the VIS/NIR/SWIR regions and records data on 224 spectral channels. These data span the electromagnetic spectrum from 0.5 μm to 2.5 μm with a wavelength resolution of approximately 10^{-2} μm. Pixels in AVIRIS scenes mostly contain background with components such as water, grass, asphalt, geological, and other material [14]. Pixels with signal present (i.e., in the plume) have a gas influencing the signal, while those without signal present do not have a gas influencing the signal. As noted in sec. 1.1, a GLS approach that matches gas library spectra has successfully indentified plume pixels by using the scene to estimate the covariance of the background [14].

This work develops a similar gas plume characterization experiment using synthetic data based on the Spatially Enhanced Broadband Array Spectrograph System (SEBASS) imaging spectrometer [36]. The SEBASS sensor was chosen because it is representative of a LWIR hyperspectral sensor, general specifications of the instrument are available (including

its impressive SNR [37]), and SEBASS data has been used for gas plume detection [38]. In addition, a LWIR instrument was chosen because data from the LWIR spectral region is less affected by reflected radiation in the scene than data collected in shorter wavelengths.

3.2 Gas Plumes

In previous chapters, the importance of a linear signal model such as the following has been emphasized:

$$\mathbf{r}_t = \mathbf{z}_t + \epsilon\mathbf{b}. \tag{3.1}$$

In this model, the observation term $\mathbf{r}_t$ is an additive combination of a gaussian background term, $\mathbf{z}_t \sim \mathcal{N}(\boldsymbol{\mu}, K)$, and a signal of interest scaled by some "strength" factor, ϵ. The SIDE filter and B-SIDE filter described in Chapter 2 assume that the data being analyzed fits this model. The CMF also assumes this model. Furthermore, the linear classifiers described in sec. 2.1 do best when the data fits that model. The classifiers that have a linear behavior can only describe straight lines (or separating hyperplanes) in the space that the data occupies. This separation capability seems sufficient for the special cases where data do follow the linear model described by (3.1). The LDA, LC, and Fisher's Discriminant all exploit this linear nature of the data. The QDA technique, which is quadratic in nature, can describe more interesting boundaries.

The data used in this work are synthetic hyperspectral images with gas plumes inserted. This data is used assuming that the signal arriving at a LWIR sensor when a plume is in the field of view in fact adheres to the the linear signal model. To show that this is so, the nature of that signal must be described.

It is important to understand that there are many paths that photons can take to arrive at a sensor pointed at the ground. There are photons emitted by materials at their respective temperatures, and transmissions through gasses, as well as absorptions and reflections by opaque objects. Though there are many paths to consider, there are only a few that significantly contribute to the signal measured at the sensor in LWIR measurements. Only those important paths will be considered in this development, which is drawn mainly from

the literature focusing on the detection problem, but also from some work done on artificial plume insertion [38–41].

Firstly, a model will be developed for paths that arrive at a sensor which is pointed at the ground, but not viewing any gaseous plumes in the atmosphere. The main contributing paths to the signal in this situation are atmospheric upwelling (L_u), background (ground) radiance (L_g), and noise (L_n). Atmospheric upwelling is radiance that comes from the atmosphere's thermal emission at its temperature. The ground radiance is described as a combination of the thermal emission of an object at the ground's temperature scaled by the emissivity of the material emitting (ε_g), and atmospheric downwelling. The noise term encompasses a few effects including the noise on the focal plane and the thermal emission of the sensor itself onto its focal plane. The combination measured at the sensor can be stated as

$$L_{clear}(\lambda) = L_u(\lambda) + L_g(\lambda)\tau_{atm}(\lambda) + L_n(\lambda), \tag{3.2}$$

which describes the attenuation of the ground radiance by the transmissivity of the atmosphere, τ_{atm}. This also denotes the wavelength (λ) dependence of all these terms. The background radiance must be further broken up into constituent terms:

$$L_g(\lambda) = B(\lambda, T_g)\varepsilon_g(\lambda) + L_d(\lambda)(1 - \varepsilon_g(\lambda)). \tag{3.3}$$

The background material is held to be at some temperature T_g, and it is assumed to be radiating as a perfect blackbody does. The $B(\lambda, T_g)$ function represents the Planck function for radiating blackbodies. This radiation is scaled by the emissivity of the material, ε_g. The other significant contribution to the background radiance term is the reflected atmospheric downwelling. The atmosphere, as previously mentioned, radiates upwards towards the sensor, but some of that energy will be radiated towards the ground and consequently reflected back towards the sensor. This radiance will not be perfectly reflected, as some will be absorbed by the material. However, for simplicity's sake, it is assumed that the particles absorbing this downwelled radiance are in local thermodynamic equilibrium [41].

That is, they are not absorbing photons and converting them to heat energy: the photons that are absorbed are completely emitted in turn. This means that $\varepsilon = \alpha$. So, in order to conserve the process, $1 = \alpha + \rho$, where ρ is the reflected portion of received radiance, and α is the absorbed portion of received radiance. By combining these two ideas, $1 = \varepsilon + \rho$. So, considering wavelength dependence and including descriptive subscripts, the reflected portion of the downwelling radiance can be expressed as

$$\rho_g(\lambda) = 1 - \varepsilon_g(\lambda), \tag{3.4}$$

which is the attenuating factor seen in (3.3) attached to the atmospheric downwelling term, L_d.

When (3.2) and (3.3) are combined, the following result is the model for the signal arriving at an airborne sensor in the LWIR when no plume is present in the field of view:

$$L_{clear} = L_u(\lambda) + [B(\lambda, T_g)\varepsilon_g(\lambda) + L_d(\lambda)(1 - \varepsilon_g(\lambda))]\tau_{atm}(\lambda) + L_n(\lambda). \tag{3.5}$$

The last assumption of local thermal equilibrium on the ground allows the radiance received at the sensor to be modeled in just a few practical terms: atmospheric upwelling and downwelling radiance (L_u, L_g), atmospheric transmittance (τ_{atm}), the emissivity of the background (ε_g), and a radiance term corresponding to noise (L_n).

The presence of the Planck function $B(\lambda, T_g)$ does not present any difficulty in the actual use of this model, as it is easily calculated using only a few known constants and the inputs shown. The Planck function can be written as

$$B(\lambda, T) = \frac{2hc^2}{\lambda^5 \left(e^{\frac{-hc}{k_B \lambda T}} - 1 \right)}. \tag{3.6}$$

The Planck function has units of Flics/μm and the constants it makes use of are the speed of light (c), Planck's constant (h), and the Boltzmann constant (k_B). The temperature of the blackbody is T, and the wavelength being measured is λ.

Including the effects of a plume in scene can be done in a few short steps. The paths that govern this model are slightly more complicated than before. The atmospheric upwelling radiance remains a significant contributor to the signal, but the downwelling radiance that is reflected to the sensor now passes through the plume. The plume has a similar effect on the background thermal radiance. The resultant effect on these signals (reflected downwelling and background thermal radiance) is further attenuation based on the transmissivity of the plume (τ_p). The radiance of the plume material itself also contributes to the signal model. This radiance (L_p) occurs at the temperature of the plume material, and must pass through (and thus be attenuated by) the intervening atmosphere to reach the sensor. The atmospheric attenuation is assumed applied to the plume at the same strength it is applied to the background thermal radiance path. This is because the plume is assumed to be close to the ground, rather than much closer to the sensor. The result of these effects can be written as

$$L_{plume}(\lambda) = L_u(\lambda) + L_g(\lambda)\tau_{atm}(\lambda)\tau_p(\lambda) + L_p(\lambda)\tau_{atm}(\lambda) + L_n(\lambda). \tag{3.7}$$

At this point it is important to point out that L_p can be expressed in terms of a Planck function and a plume material emissivity:

$$L_p(\lambda) = B(\lambda, T_p)\varepsilon_p(\lambda), \tag{3.8}$$

where $\varepsilon_p(\lambda)$ is the emissivity of the plume material. This expression of L_p serves to make the expression of more practical use by breaking down the thermal radiance term into a measurable emissivity and a calculable Planck function.

Using (3.3) to expand L_g in (3.7), the model for radiance received with a plume in view is

$$\begin{aligned} L_{plume}(\lambda) &= L_u(\lambda) + [B(\lambda, T_g)\varepsilon_g(\lambda) + L_d(\lambda)(1 - \varepsilon_g(\lambda))]\tau_{atm}(\lambda)\tau_p(\lambda) \\ &\quad + L_p(\lambda)\tau_{atm}(\lambda) + L_n(\lambda). \end{aligned} \tag{3.9}$$

This statement describes many of the important paths and their terms, but what is desired is to express this model with a signal term (and an associated strength) and an additive noise term. The signal term should also be a function of the signature of the chemical being detected, $b(\lambda)$, which does not appear in (3.9). The next steps are made in order to arrive at a linear expression from this model that explicitly involves $b(\lambda)$.

The transmissivity of the plume can be expressed as follows:

$$\tau_p(\lambda) = \exp(-cb(\lambda)) \approx 1 - cb(\lambda), \tag{3.10}$$

where c is the column density of the gas plume, and the chemical signature of the gas is $b(\lambda)$. This transmissivity is approximately linearized using the assumption that $cb(\lambda)$ is small. The assumption of a small $cb(\lambda)$ must be kept true, otherwise the linear model will be unfaithful to the physical process. This is assumed to be a fair enough assumption when c is small, which is what is meant by a gas plume being optically "thin."

The plume emissivity, which is the scaling factor in (3.8), can be related to $\tau_p(\lambda)$ and therefore to $b(\lambda)$ in this way:

$$\varepsilon_p(\lambda) = 1 - \tau_p(\lambda) \approx cb(\lambda). \tag{3.11}$$

This relationship allows the full radiance model in (3.9) to be expressed linearly in $b(\lambda)$. Substituting (3.10) and (3.11) into (3.9), the model can now be written

$$\begin{aligned} L_{plume}(\lambda) &= L_u(\lambda) + [B(\lambda, T_g)\varepsilon_g(\lambda) + L_d(\lambda)(1 - \varepsilon_g(\lambda))]\tau_{atm}(\lambda)(1 - cb(\lambda)) \\ &\quad + cb(\lambda)B(\lambda, T_p)\tau_{atm}(\lambda) + L_n(\lambda) \\ &= L_u(\lambda) + [B(\lambda, T_g)\varepsilon_g(\lambda) + L_d(\lambda)(1 - \varepsilon_g(\lambda))]\tau_{atm} + L_n(\lambda) \\ &\quad + cb(\lambda)\left[B(\lambda, T_p) - (B(\lambda, T_g)\varepsilon_g(\lambda) + L_d(\lambda)(1 - \varepsilon_g(\lambda)))\right]\tau_{atm}(\lambda). \end{aligned} \tag{3.12}$$

A linear model with an additive noise term can now be realized using the result in (3.12). By allowing the background, or noise term to be

$$n(\lambda) = L_u(\lambda) + [B(\lambda, T_g)\varepsilon_g(\lambda) + L_d(\lambda)(1 - \varepsilon_g(\lambda))]\tau_{atm} + L_n(\lambda), \tag{3.13}$$

and letting the signal term be $b(\lambda)$ and its associated strength term be

$$\epsilon(\lambda) = c\left[B(\lambda, T_p) - (B(\lambda, T_g)\varepsilon_g(\lambda) + L_d(\lambda)(1 - \varepsilon_g(\lambda)))\right]\tau_{atm}(\lambda), \tag{3.14}$$

the model can expressed linearly as

$$L_{plume}(\lambda) = \epsilon(\lambda)b(\lambda) + n(\lambda). \tag{3.15}$$

These can then be expressed in vector notation by letting each entry indexed by i in the vectors correspond to a λ_i that is significant because of a sensor being used or some other practical motivation. Then let

$$\begin{aligned} \mathbf{L}_{plume} &= \mathbf{L}_u + [\mathbf{B}(T_g) \odot \boldsymbol{\varepsilon}_g + \mathbf{L}_d \odot (1 - \boldsymbol{\varepsilon}_g)] \odot \boldsymbol{\tau}_{atm} + \mathbf{L}_n \\ &\quad + c\mathbf{b} \odot [\mathbf{B}(T_p) - (\mathbf{B}(T_g) \odot \boldsymbol{\varepsilon}_g + \mathbf{L}_d \odot (1 - \boldsymbol{\varepsilon}_g))] \odot \boldsymbol{\tau}_{atm}, \end{aligned} \tag{3.16}$$

and let the linearization assignments be made for these terms in the same way they were made for the wavelength-continuous expressions. These values can now be assigned to the vector-valued variables in (3.1):

$$\mathbf{z}_t = \mathbf{L}_u + [\mathbf{B}(T_g) \odot \boldsymbol{\varepsilon}_g + \mathbf{L}_d \odot (1 - \boldsymbol{\varepsilon}_g)] \odot \boldsymbol{\tau}_{atm} + \mathbf{L}_n, \tag{3.17}$$

and in this case, the signal strength term is actually expressed as an element-by-element multiply where $\boldsymbol{\epsilon} \odot \mathbf{b}$ results in the signal strength being applied. The value of this strength term is

$$\boldsymbol{\epsilon} = c[\mathbf{B}(T_p) - (\mathbf{B}(T_g) \odot \boldsymbol{\varepsilon}_g + \mathbf{L}_d \odot (1 - \boldsymbol{\varepsilon}_g))] \odot \boldsymbol{\tau}_{atm}. \tag{3.18}$$

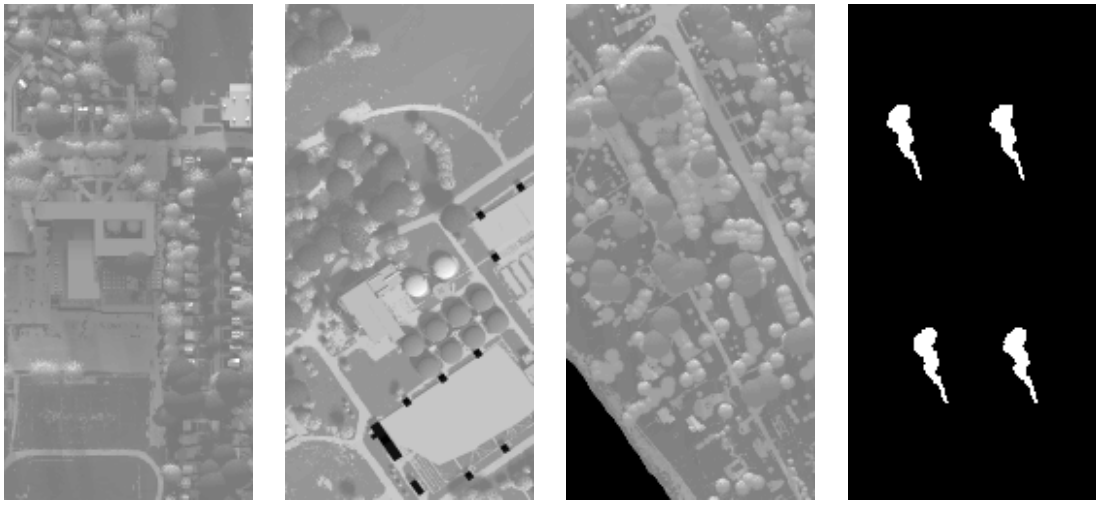
For the purposes of this work, however, the element-by-element scaling is assumed to have been applied to the known spectrum $\mathbf{b}$ as a pre-processing step, so only the column density c remains as a signal strength scaling term (ϵ).

The linear model (3.1) reiterated at the beginning of this section can now be constructed using these assignments. This allows the processing techniques based on these linear models to be applied with great success. This development shows that the linear model is a valuable, and effective approximation to the radiometric models. It is also clear that when assumptions about column density being small or the plume being close to the ground are broken, this model may cause algorithm performance to degrade.

There are also discussions that consider the below-plume atmospheric upwelling, and above-plume atmospheric upwelling as separate terms in the case that the plume is not close to the ground compared to its distance to the sensor [40]. The plumes studied for the experiments in this work are assumed to be near enough to the ground that the difference is negligible.

3.3 DIRSIG

To demonstrate the SIDE technique, synthetic hyperspectral data was generated to represent sensor measurement of a realistic scene that included gas plumes (i.e., the signal of interest). The synthetic data is used to allow exact comparisons between the exploited data and the actual information. These synthetic data were generated using the DIRSIG program, from the Digital Imaging and Remote Sensing (DIRS) Lab at the Rochester Institute of Technology [1]. DIRSIG generates synthetic imagery for a specific physical scene based on the scene materials, atmospheric environment, and specific sensor used for measurement. DIRSIG includes a detailed model of the sensor spectral response. For the experiments in this work, the sensor was based on the long-wave channels of the SEBASS sensor using a pushbroom operation for data collection [36]. The SEBASS description used has a spectral response with 128 spectral bands centered from $7.45 - 13.2\,\mu$m with 128 elements across the focal plane. The images were captured 256 lines at a time, for a total of 32 768 pixels in a data cube ($128 \times 128 \times 256$). DIRSIG uses the MODTRAN code to



(a) (b) (c) (d)

Fig. 3.1: (a), (b), (c) Broadband images of the three scenes used in the data set. The large black portion of (c) is water. (d) The four locations where plumes were inserted (one at a time in each image) to diversify the background.

incorporate atmospheric radiative transfer models into the imagery [42].

This data generation step takes advantage of a high resolution physical scene model of areas in Rochester, New York supplied with DIRSIG to provide image background at the collection site. Three different physical locations were chosen in the Rochester model for this experiment. The locations chosen are a residential area, with many small buildings as shown in fig. 3.1(a), an industrial area featuring large metal structures and asphalt as shown in fig. 3.1(b), and a lakeside residential area shown in fig. 3.1(c).

DIRSIG documentation recommends generating the image data at a spatially oversampled rate of 4×4. This means that each pixel in the resulting data cubes is the average of a 4×4 square of pixels from the original high-resolution image. This oversampling and averaging improves the reality of the data.

The synthetic data cubes were augmented with a phenyl benzene plume. The plumes were generated using the Blackadar plume generation model [43]. The Blackadar model computes plume generation data using a "drunkard's walk" random approach. For these plumes a gas release temperature of $350\,K$ is assumed, with an ambient air temperature of $293\,K$ and winds of 4.0 m/s, 290° from North for the release. The plume was created in

16 steps with 16 puffs of plume per time step. The puffs cool and diffuse as time passes according to the Blackadar model.

This approach provides synthetic data cubes based on the SEBASS sensor for exploitation and ground truth maps with the actual plume locations as generated by DIRSIG. These ground truth maps also contain information about material and temperature which affect gas detector performance, but these data were not used in this experiment. Figure 3.1(d) is a composite of four of these maps, showing the four release locations where plumes were inserted into the images. Only one plume at a time was inserted in each data cube with a number of data cubes generated for each scene.

Chapter 4

Results

The experiments done for this study included generating data and associated ground truth, running the various algorithms on this data, and quantifying the results in ROC curves and calculating the area under these curves for each algorithm test. These tests were run while varying a number of variables that are described in the first section of this chapter. The results are presented so that an easy comparison to the existing techniques might be made. First, results from testing the SIDE filter are shown with plots of ROC curves and a few other average results. These are compared to the CMF for a standard baseline to measure against. The SIDE filter's effectiveness is clear in these experiments. Secondly, the B-SIDE filter is compared to the CMF as well as the statistical techniques from sec. 2.1. The B-SIDE filter is proposed as a post-processing technique that can be applied to any initial detection algorithm output. The results show that the B-SIDE filter is a valuable tool to add to the group of trained classifying techniques.

This chapter accomplishes a summary of the results and commentary on these new filters in the context of older techniques. The full results can be found in the appendices (A and B). Plots showing every average ROC curve can be found there along with tables of the average area under these curves for all experiments.

4.1 Experimental Data Set

This section will explain the process followed in generating the synthetic data used in most of the work described in this document. The DIRSIG Model which was used to generate this data is based on first principles. This means that the radiometry is as near to real as can be simulated. It was utilized to generate a set of 192 images that were applied in various tests to evaluate SIDE and its derivatives.

Using DIRSIG, we generated hyperspectral data cubes for each combination of: three scenes, four plume release rates, four release locations, two atmospheres, and two sensor elevations resulting in 192 data cubes for the study. We chose the scene and release location parameters to represent a variety of conditions to reduce the dependence of the results on scene content, plume background, and plume location in scene. The following results are average performance results at various release rates, atmospheres, and sensor elevations, as well as 10 different additive noise realizations per data cube.

The SIDE and experiment variables studied were SNR, α, and ϵ. Two values (60dB and 80dB) were chosen for the SNR which was calculated as $SNR \equiv 10\log_{10}(E[\mathbf{x}^T\mathbf{x}]/E[\mathbf{n}^T\mathbf{n}])$ where $\mathbf{x}$ is zero-mean and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$ is additive noise [44]. Values were also chosen for $\alpha \in \{10^0, 10^1, 10^2, 10^3\}$ and for $\epsilon \in \{10^{-5}, 10^0, 10^2\}$ in order to observe behavior over large ranges of values.

DIRSIG is now distributed with a graphical user interface that greatly reduces the text editing requirements in setting up a simulation to generate image data. The interface allows the user to specify atmospheric conditions, sensor specifications, platform motion, etc. Of great importance to these experiments was the ability to insert chemical plumes into the images. That capability is not currently integrated into the graphical user interface. This required the `dirsig` command to be executed from a command prompt. In addition to this requirement, the fact that it was desired to automatically generate many images with different characteristics motivated the "old-fashioned" approach of text-editing.

For each image that is generated using DIRSIG, a list of files describing the simulation must be generated. The basic files needed are a

- simulation file,
- scene file,
- atmosphere file,
- platform (sensor) file,
- platform motion file,

- and a task file.

The simulation file provides DIRSIG a list of all the other files that are needed for the simulation. The scene file describes the scene to be used in the simulation, and if a plume is going to be inserted, describes the plume parameters. This file allows the scene geometry and material contents to be accessed for the simulation. The atmosphere file helps describe what atmosphere is used in the simulation. Atmosphere effects are calculated using a separate program called MODTRAN. The platform file describes the sensor. This file tells the simulation what the sensor looks like, including the size and shape of the focal plane (or planes) and its elements, its focal length, and provides collection information such as a list of what information to output in the truth maps. This file also describes how highly the data is oversampled and can specify a point spread function to apply to the collection. The platform motion file describes the physical location of the sensor relative to the scene, and also its motion relative to the scene. An altitude and velocity can be specified in this file. The task file describes when this simulation will record data, relative to the motion sequence described in the platform motion file. This file allows a complicated motion sequence to be described one time in the motion file, but separate or partial collections to be performed.

Most of these files required a small amount of editing for each image that was generated. This was accomplished with short Perl scripts that changed file names or numerical values, such as focal lengths or plume release rates. After the appropriate editing was done and the DIRSIG simulation was completed, an output of a hyperspectral image and associated truth data in an ENVI format was saved.

4.2 SIDE

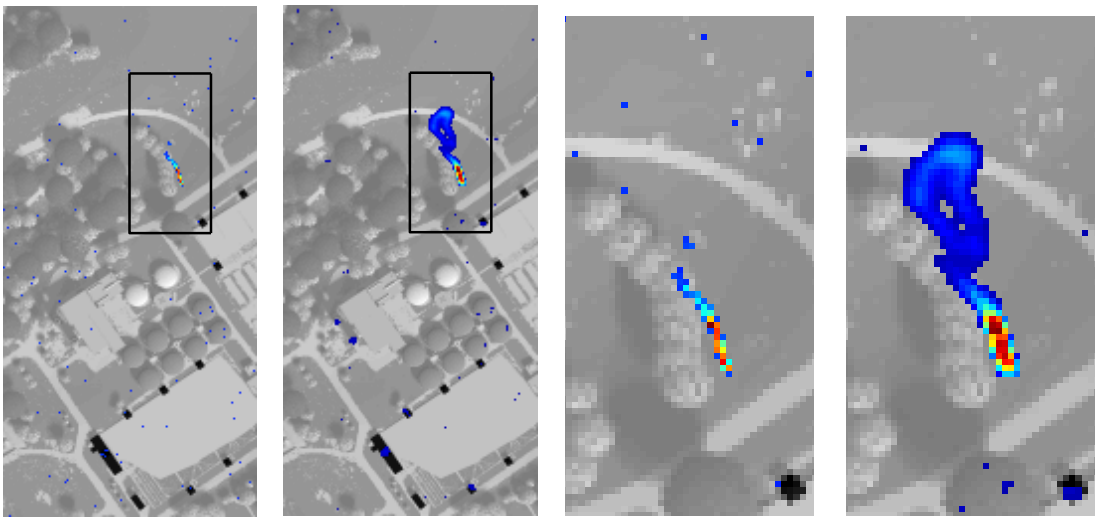
The SIDE filter results were generated using the experimental description in the previous section. This section of results shows only results from using the vector-valued SIDE filter compared to the CMF. The results in this section were mostly generated using a 3×3 neighborhood for the SIDE filter, but a short summary of results from tests using the 13-pixel neighborhood are also included in this section.

4.2.1 Comparison to CMF

Gas plume detection using CMF is compared to using CMF with SIDE post-processing. Figure 4.1 shows the CMF output on the left (figs. 4.1(a) and 4.1(c)) and the results of the SIDE filter on the right (figs. 4.1(b) and 4.1(d)). A detection threshold for these images was chosen that yielded a $P_{FA} = 0.5\%$. Figure 4.1 shows pixels greater than this threshold overlain in color on a grayscale background image of the scene. The plume colors represent signal strength which can be related to the total gas path length or approximate concentration. These images are from a cube with a mid-latitude winter atmosphere, the sensor 5 km above the ground, and a plume concentration of 500 ppm. The SIDE filter used $\alpha = 100$ and $\epsilon = 1$. Visual inspection shows that the SIDE filter identifies large parts of the plume that were not detected by the CMF algorithm. In addition, it appears that the SIDE filter suppresses some of the false positives related to noise and slightly enhances some false-positive anomalies in the CMF image that most likely resulted from materials that had similar spectral features to phenyl benzene. Despite that enhancement of non-noise related anomalies, the SIDE filter vastly improves detection for the fixed P_{FA}.

The plume release (and by inference plume concentration) is the most important performance measure for evaluating the CMF and SIDE filters. The release rate is directly related to the optical thickness of the plume, especially under synthetic conditions where all other variables are constant. For the synthetic data, the release rate is directly proportional to the signal strength being detected. The experiments showed that generally, the SIDE filter dramatically increases detection levels (i.e., is able to detect weaker signals) without increasing the false-positive detections alarms at all release rates. These tests showed that the SIDE filter performed better on the lowest release rate than the CMF alone did on the highest release rates used in this experiment.

The results from these tests are compared with the ground truth maps to generate receiver-operating characteristic (ROC) curves for each detector to quantitatively compare the two approaches. The ROC curves are presented in fig. 4.2. In these figures, dashed curves and solid curves represent the CMF and SIDE detectors, respectively. In fig. 4.2(a),



(a) (b) (c) (d)

Fig. 4.1: A gray broadband image is the background for these example images of the scenes where plumes have been inserted. (a) The CMF detects a portion of the signal when P_{FA} is fixed at 0.5%. (b) The SIDE filter detects portions of the signal not detected by the CMF with the same P_{FA}. In (c) a close-up of the area in the box in (a) is shown, and (d) shows a similar close-up for the SIDE filter image in (b).

the line markers correspond to the release rates for each experiment. Figure 4.2(a) shows that the plume with the highest release rate did not provide the best in detection performance. Our model assumes optically-thin plumes such that the plume presence is additive to the background image and does not obscure it. This assumption starts to be violated as the plume becomes optically-thick, obscuring the background. When the plume becomes optically-thick, the optical path length becomes large enough that the assumptions made in linearizing the model describing the plume presence are violated [35]. The CMF performance consequently degrades, also influencing the SIDE filter performance as it is a post-processing filter. Even with these issues, fig. 4.2(a) shows that the SIDE filter improves detection capability at all four release rates tested.

In fig. 4.2(b) the line markers represent two different atmospheres. This shows the impact of the atmospheric column on detector performance. We chose a summer and winter realization of a mid-latitude atmosphere meant to model the conditions typical to those in Rochester. Figure 4.2(b) shows how the different atmospheres affect detection using the

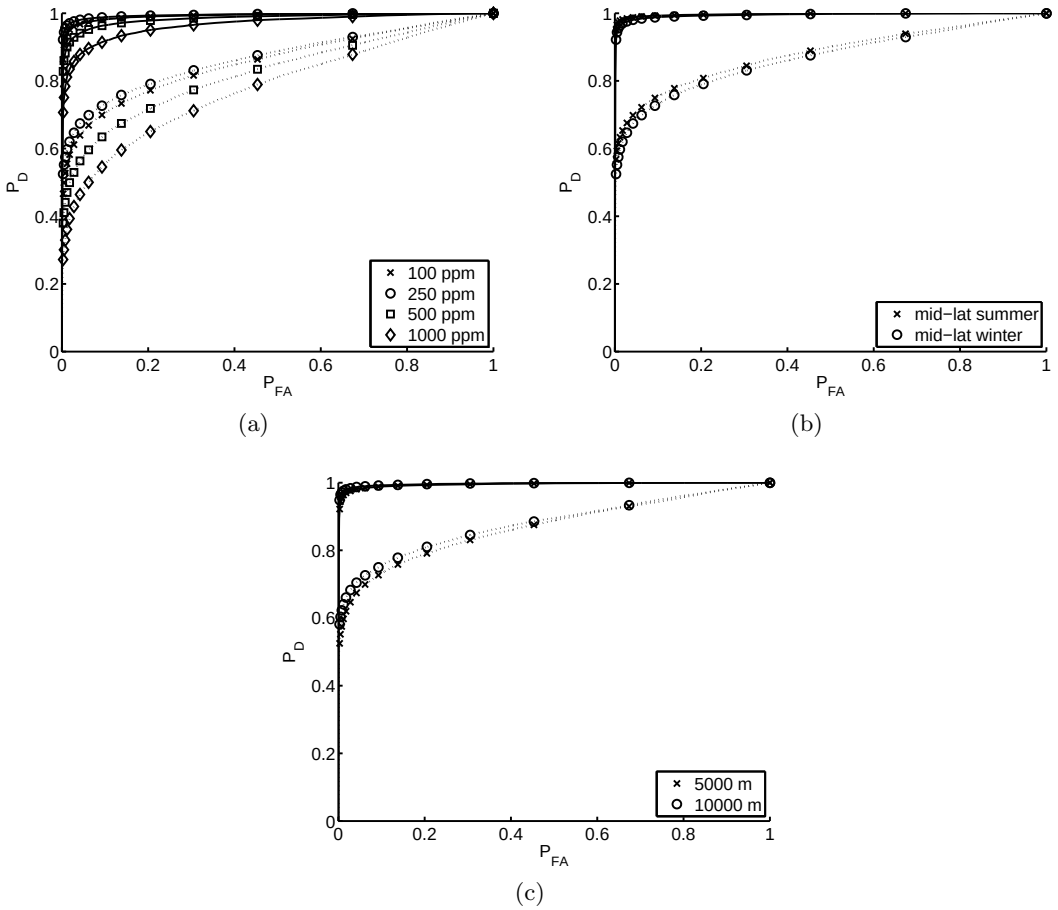


Fig. 4.2: (a) Performance of the SIDE filter compared to the CMF when plumes of different release rates are detected. (b) SIDE versus CMF performance in two different atmospheres, one each from a different time of the year collected at Rochester, New York. (c) SIDE versus CMF performance at two different sensor elevations. The ROC curves corresponding to the performance of the SIDE filter are solid lines, while the dashed lines correspond to CMF performance.

CMF algorithm. As expected, this also affects the SIDE filter. However, the difference between SIDE filter results is slightly smaller than the difference in CMF results, nearly overcoming the effects of time of year on detection. As with release rates, the SIDE filter significantly improves the detector performance.

Figure 4.2(c) shows the impact of sensor elevation on detector performance. The sensor was placed at 5 km and at 10 km to study the effect of increasing atmosphere on detection capability. The atmosphere represents signal degradation in sensor channels and attenuates

certain frequencies more than others. The effect on CMF performance with the sensor at 5 and 10 km is similar to the difference between the summer (high humidity) and winter (colder and dryer) atmospheres. As with the difference in atmospheres, the altitude difference affects the SIDE filter less than it affects the CMF, showing the SIDE filter robustness to these variables.

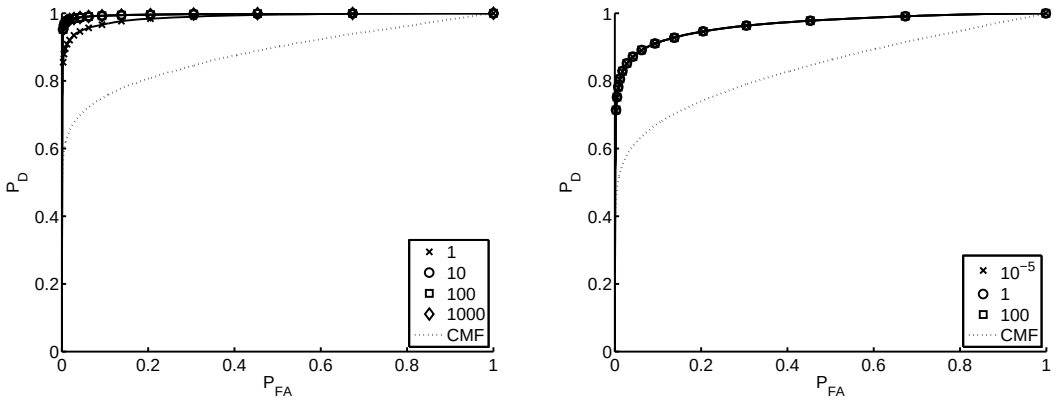
4.2.2 SIDE Filter Parameters

In addition to comparing the CMF and SIDE filters, the impact of the probability model parameters were explored. Figure 4.3(a) shows the results of varying the α parameter. This determines the emphasis on signal contiguity. Larger α values represent more emphasis on contiguity. As fig. 4.3(a) shows, larger values result in better performance because of the contiguity in the plume in the experimental data set. Figure 4.3(b) shows the impact of changing ϵ on the SIDE filter. Figure 4.3(b) shows that the SIDE filter is insensitive to changes across a wide range of ϵ and is robust to choice of ϵ. Theiler et al. propose that when the background is assumed to be gaussian and the signal additive, then the CMF does not depend on the signal strength [31]. The results of this experiment seem to suggest that since these same assumptions were made in the development of the SIDE filter (see (2.41)), it maintains the same invariance to signal strength.

Table 4.1 presents a quantitative comparison of the performance of the CMF and SIDE filters. In Table 4.1, A_C and A_S represent the average areas under the ROC curve for the CMF and SIDE filters, respectively. The difference in performance is denoted as $\Delta A = A_S - A_C$. Table 4.1 presents results for data cubes generated using a mid-latitude winter atmosphere and a sensor at 10 km. The neighborhood processing was done with $\alpha = 1000$ and $\epsilon = 1$. The table also presents results with two different noise levels in the data. The table clearly shows an increase in detector performance for the SIDE filter.

4.2.3 13-Pixel Neighborhood

A less extensive test was run using the neighborhood from fig. 2.3(a) on the same data used in the tests using the 3×3 neighborhood. The raster scan method for calculating



(a) (b)

Fig. 4.3: (a) The effect of α from (2.42) on SIDE filter performance, and (b) performance of the SIDE filter for different values of ϵ with $\alpha = 1$. These curves were obtained from a data cube with a plume release rate = 250 ppm.

Table 4.1: Area under ROC curve for SIDE and CMF at mid-lat winter and sensor at 10 km.

	SNR (dB)					
	50			60		
rate (ppm)	A_C	A_S	ΔA	A_C	A_S	ΔA
100	0.7242	0.92205	0.19785	0.85245	0.99194	0.13949
250	0.76873	0.9558	0.18707	0.87959	0.99632	0.11673
500	0.81484	0.97948	0.16464	0.86284	0.99374	0.1309
1000	0.8186	0.9824	0.16381	0.84348	0.9903	0.14683

$\Sigma\Delta$ was altered slightly to fit the different shape of this neighborhood (fig. 2.3(b)). This experiment reveals the cost of using larger neighborhoods: because of the larger size of the neighborhood, the sums over $\mathbf{s}$ in (2.50) has 2^4 times more terms than the one for the 3×3 neighborhood. This means a nearly $16\times$ increase in computations. Extending to a 5×5 neighborhood would cause the sum to have 2^{16} times more terms, which is much more prohibitive than a $16\times$ increase in computation requirements.

These results were calculated using images generated with a mid-lat summer atmosphere and a sensor at 10 km. The plumes tested had release rates of 100 ppm and 500 ppm, and the tests were done at 50 dB and 60 dB SNR. The results in fig. 4.4 show results from this larger neighborhood compared to the 3×3 neighborhood and the CMF. These

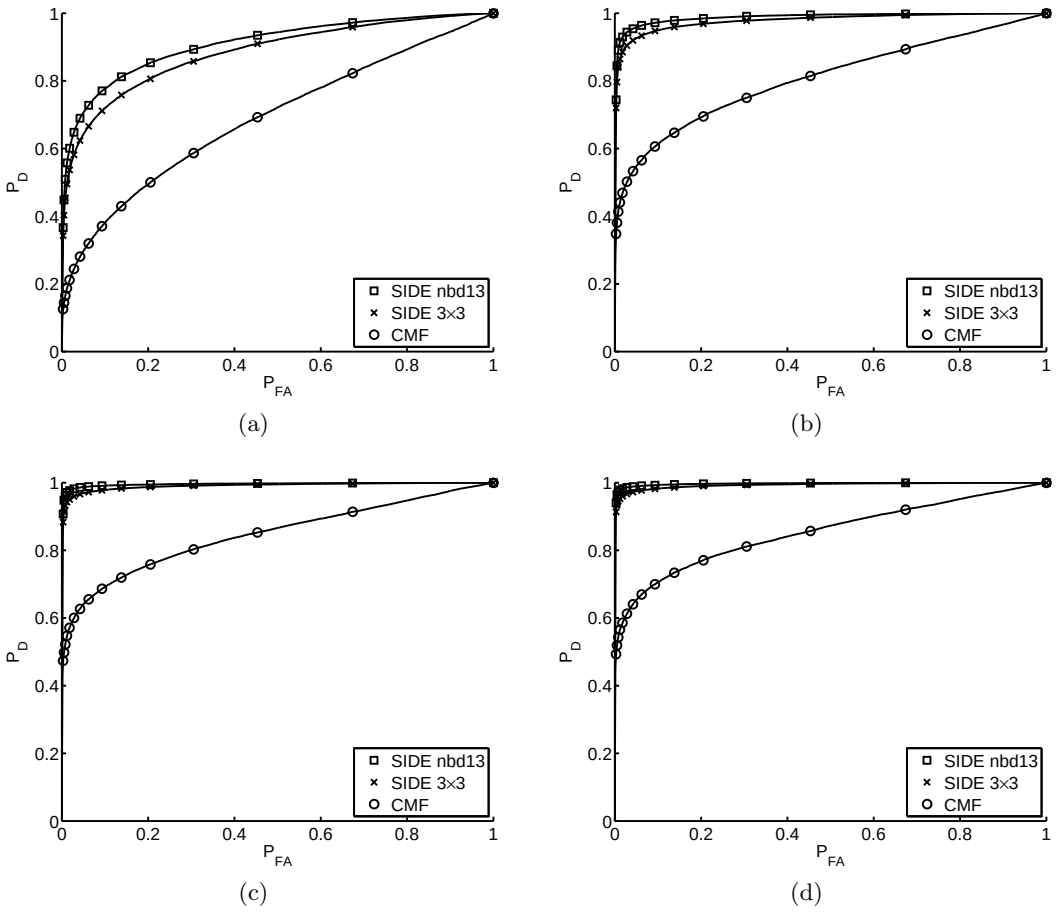


Fig. 4.4: SIDE filter with two different neighborhood sizes and CMF at mid-lat summer, sensor at 10 km. In (a), release rate = 100 ppm and SNR = 50 dB, (b) release rate = 100 ppm and SNR = 60 dB, (c) relase rate = 500 ppm and SNR = 50 dB, and (d) release rate = 500 ppm and SNR = 60 dB.

results show that the 13-pixel neighborhood out-performed the 3×3 at both rates and noise levels tested. Table 4.2 shows the area under the ROC curve for each of the curves in fig. 4.4. In this table, $A_{S_{3\times3}}$ denotes the area under the ROC curve for the SIDE filter performance using a 3×3 neighborhood size, and $A_{S_{13}}$ denotes the same for the SIDE filter using a 13-pixel neighborhood. The CMF performance is denoted by A_C.

4.3 B-SIDE

The B-SIDE filter developed in sec. 2.3.2 was applied to the data set used for evaluating the SIDE filter. The goal of this application was to use the B-SIDE filter in a way that

Table 4.2: Area under ROC curve for CMF and SIDE using 3×3 neighborhood and 13–pixel neighborhood at mid-lat summer and sensor at 10 km.

	SNR (dB)					
	50			60		
rate (ppm)	A_C	$A_{S_{3\times3}}$	$A_{S_{13}}$	A_C	$A_{S_{3\times3}}$	$A_{S_{13}}$
100	0.7242	0.8834	0.9094	0.85245	0.9782	0.9881
500	0.81484	0.9904	0.9950	0.86284	0.9925	0.9964

avoids the human in the processing loop as much as possible. It is assumed that the work to prepare the chemical signature for matching is done according to the model developed in sec. 3.2, given that the signature has been collected in a laboratory setting and other processing has been done specific to the image under test. In the case of other detectors (not the CMF), the output image is still desired to be further enhanced without the aid of a person. The only obstacle to making this an automatic step is that the B-SIDE filter requires a target class mean and a background class mean as inputs. These means must be calculated in some way. The options at this point in the processing include: (1) present training data to the B-SIDE filter, allow it to calculate class means based on that training data, and then do the SIDE filtering to enhance the detection image; and (2) use a clustering algorithm to calculate means and use those clustered means as the inputs to the B-SIDE filter. This second approach was was chosen for test, since it holds more potential for real-world applications.

4.3.1 Comparison to CMF

In order to test the real viability of this B-SIDE configuration of the SIDE filter, the CMF was used as a the scalar input. Two class means were calculated from this CMF scalar image, and subsequently used as the parameters for the B-SIDE filter. As stated in the previous paragraph, the means were calculated using a clustering algorithm. For these experiments, a k-means algorithm was used to find two different class means. This algorithm was set to iterate a maximum of 100 times before stopping. At this point, the minimum of the two means was used for the background class mean, and the maximum of

the two was used as the target class mean. A block diagram showing this process is shown in fig. 4.5.

As was the case in the SIDE filter results in fig. 4.2(a), the B-SIDE filter with the CMF as input improved the results dramatically. When compared with the plume release rate, the B-SIDE filter processing CMF output, which we will write as B-SIDE + CMF, improves the worst performing CMF to better than the best performing CMF alone (fig. 4.6). These results are very similar to the results for the SIDE filter in its original configuration. This seems to show that the B-SIDE filter is an effective bolt-on tool, performing as well as a post-processing step as it does in the original vector-valued configuation.

The data shown in Table 4.3 show that the improvement is better when B-SIDE is used in circumstances that yield poorer performance in the CMF. The columns in Table 4.3 are labeled much the same way as in Table 4.1, except that instead of A_S, the column labeled A_{B_C} denotes the B-SIDE + CMF output.

4.3.2 Comparison to Statistical Classifiers

The B-SIDE filter was also applied to the output images of the statistical classifiers developed in Chapter 2. The same k-means algorithm was used to estimate class means for the B-SIDE filter as shown in fig. 4.5. The same experiments were done for these as the B-SIDE + CMF combination.

The statistical classifiers were trained on a randomly selected 80% of the current image, and then tested on the remaining 20% percent of the data. This was done so that the trained classifier could be tested, but not across other variables that might affect the performance

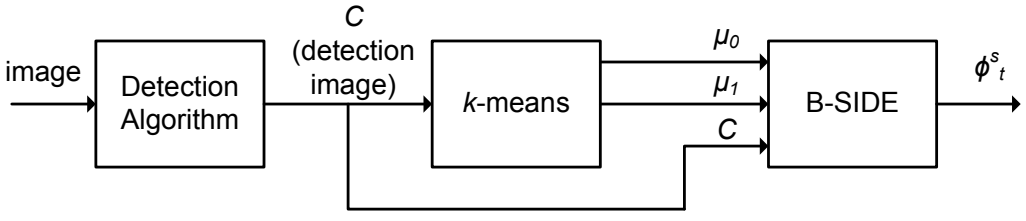


Fig. 4.5: The B-SIDE filter as a post-processing accessory, showing k-means clustering used to estimate the class means used in B-SIDE filtering.

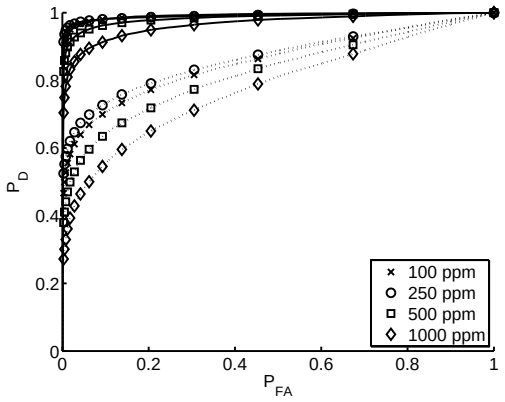


Fig. 4.6: The CMF filter as a scalar input to the B-SIDE filter.

Table 4.3: Area under ROC curve for CMF and B–SIDE + CMF at mid-lat winter and sensor at 10 km.

	SNR (dB)					
	50			60		
rate (ppm)	A_C	A_{B_C}	ΔA	A_C	A_{B_C}	ΔA
100	0.7242	0.91998	0.19578	0.85245	0.99169	0.13924
250	0.76873	0.95458	0.18584	0.87959	0.99621	0.11662
500	0.81484	0.97885	0.16401	0.86284	0.99356	0.13072
1000	0.8186	0.98185	0.16326	0.84348	0.99003	0.14655

of the classifier, such as atmosphere or sensor elevation. The ROC curves shown in fig. 4.7 were generated using only the 20% of the data that was not used for training. In the testing of the B-SIDE filter, the ROC curves were generated using only that same 20% of data. The results shown in fig. 4.7 all correspond to tests done on cubes generated with a mid-lat winter atmosphere and the sensor at 10 km, and an $\alpha = 1000$. Note that the curves corresponding the baseline technique in fig. 4.7 are dashed lines, while the curves corresponding the B-SIDE post-processing are solid lines. Take special note when referring to fig. 4.7 that the limits on the axes of the four plots shown are not the same. These limits were chosen for each statistical classifier to best show the separation in performance and curve shape. These do not correspond visually to the plots shown for the SIDE filter in the previous sections of this chapter.

In fig. 4.7(a), the results of a test on B-SIDE + LDA show the improvement over LDA

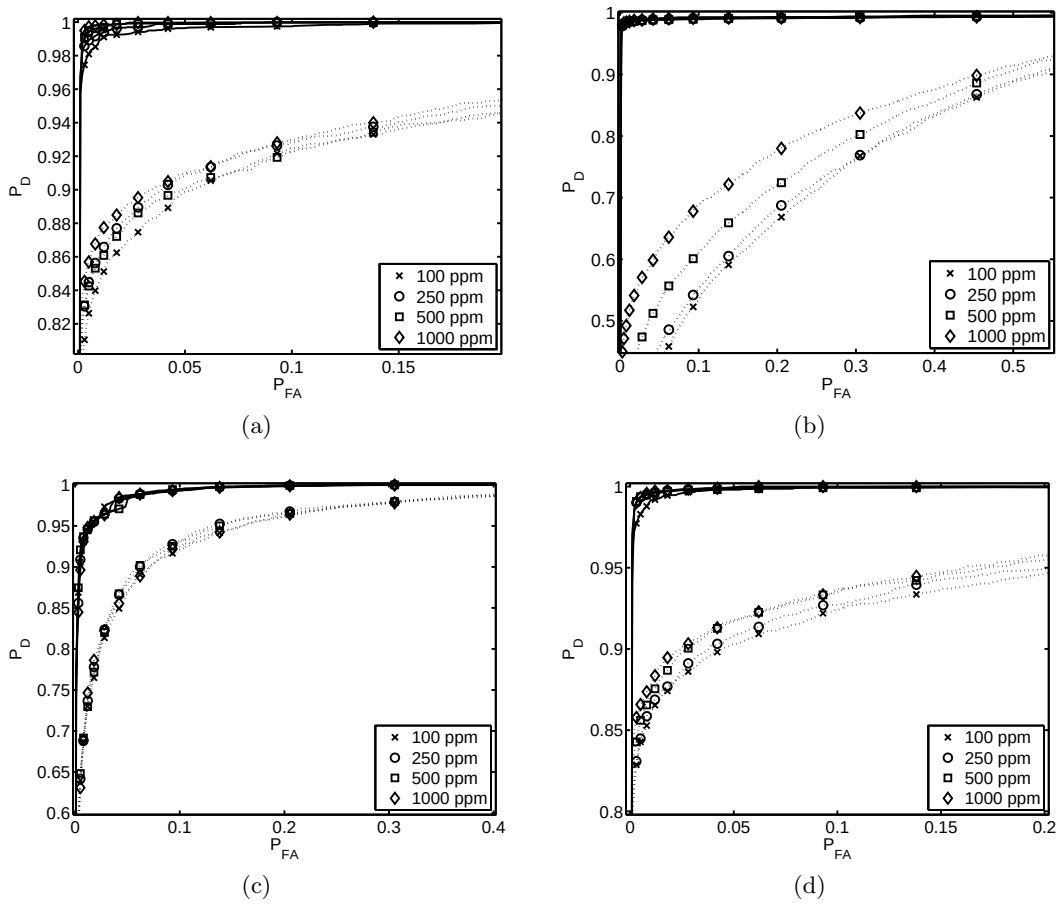


Fig. 4.7: Statistical classifiers as scalar input to the B-SIDE filter: (a) LDA, (b) QDA, (c) LC, (d) Fisher's discriminant. The dashed lines are baseline performance, and the solid lines are B-SIDE post-processing performance. Note that the limits on the axes are not uniform across these four plots.

performance at four plume release rates. When training data is available, classification can be spectacularly effective, but the post-processing done by B-SIDE can extract even more information. The LDA technique performs very well on its own in this synthetic data, which fits the linear model very well. The B-SIDE performance leaves almost nothing to be desired, however, as the area under the ROC curve for these tests is very near to 1, as shown in Table 4.4.

Figure 4.7(b) shows that the QDA technique does not perform as well as the LDA on these synthetic data. This perhaps shows that the QDA technique was overfitting this synthetic data and as a result misclassifying some of the testing data. This might have

Table 4.4: Area under ROC curve for LDA and B–SIDE + LDA at mid-lat winter and sensor at 10 km.

	SNR (dB)					
	50			60		
rate (ppm)	A_{LDA}	$A_{B_{LDA}}$	ΔA	A_{LDA}	$A_{B_{LDA}}$	ΔA
100	0.96667	0.99875	0.032084	0.98802	0.99946	0.011442
250	0.97099	0.99923	0.028244	0.9918	0.99948	0.0076795
500	0.9697	0.99928	0.029583	0.99042	0.99948	0.009058
1000	0.97299	0.99939	0.026392	0.9896	0.99947	0.0098707

been remedied by changing the ratio of training data to test data, or changing SNR. What is remarkable, however, is that the B-SIDE filter overcomes these performance issues very well. Note the nearly uniform performance across release rate in the B-SIDE + QDA curves when compared to the QDA alone curves. Table 4.5 shows data which supports the conclusion that B-SIDE filtering on QDA data helps overcome release rate issues as well as the overfitting to which QDA is prone.

The LC technique is a slightly different classifier than the previously discussed discriminant analyses. Figure 4.7(c) shows the results for this mid-lat winter, sensor at 10 km configuration. Note that the LC seems to perform better than QDA on its own, but the B-SIDE filter as a post-processor is unable to overcome its shortcomings as it does in the QDA case. The corresponding data in Table 4.6 reveal the improvement is slightly less dramatic for LC than QDA, and nearly uniform across relase rate.

Fisher's discriminant, which is very closely related to the LDA, performs fairly well alone, and can be improved very well by the B-SIDE filter. The curves in fig. 4.7(d) show a similar situation to that of LDA as far as performance without the B-SIDE filter, and improvement due to post-processing with the B-SIDE filter. Fisher's discriminant is helpful in discriminating the two classes, and lends itself well to the post-processing use of B-SIDE. Table 4.7 shows that performance increases are best at lower SNR, but the higher SNR allows nearly perfect performance.

Table 4.5: Area under ROC curve for QDA and B–SIDE + QDA at mid-lat winter and sensor at 10 km.

	SNR (dB)					
	50			60		
rate (ppm)	A_{QDA}	$A_{B_{QDA}}$	ΔA	A_{QDA}	$A_{B_{QDA}}$	ΔA
100	0.82027	0.99504	0.17477	0.94388	0.99424	0.050354
250	0.8268	0.99357	0.16677	0.94836	0.99308	0.044718
500	0.85007	0.99385	0.14378	0.95385	0.9941	0.040255
1000	0.87454	0.99382	0.11928	0.95429	0.993	0.038706

Table 4.6: Area under ROC curve for LC and B–SIDE + LC at mid-lat winter and sensor at 10 km.

	SNR (dB)					
	50			60		
rate (ppm)	A_{LC}	$A_{B_{LC}}$	ΔA	A_{LC}	$A_{B_{LC}}$	ΔA
100	0.95143	0.99283	0.041403	0.9714	0.9963	0.024899
250	0.95569	0.99448	0.038795	0.97278	0.99639	0.023614
500	0.95604	0.99494	0.038904	0.97226	0.99631	0.024052
1000	0.95442	0.99403	0.03961	0.97089	0.99617	0.02528

Table 4.7: Area under ROC curve for Fisher's discriminant and B–SIDE + Fisher's discriminant at mid-lat winter and sensor at 10 km.

	SNR (dB)					
	50			60		
rate (ppm)	A_{FISH}	$A_{B_{FISH}}$	ΔA	A_{FISH}	$A_{B_{FISH}}$	ΔA
100	0.96734	0.99912	0.031778	0.98771	0.99947	0.011762
250	0.97164	0.99927	0.027627	0.98948	0.99947	0.0099901
500	0.97361	0.99906	0.025457	0.99137	0.99948	0.0081094
1000	0.97288	0.99932	0.026442	0.98921	0.99947	0.010263

Chapter 5

Conclusions and Future Work

The contribution of this thesis is to introduce the value of the SIDE filter and the processing framework described in developing it. The SIDE filter can be very useful in extracting through a probabilistic means the spatial information surrounding an observation. This framework has been discussed at length in various settings, including vector observations and scalar observations. The framework heavily depends on the assignment of meaningful prior probabilities to the information extracted from the neighborhoods of observations. These prior probabilities have been explored briefly.

The SIDE filter has been shown to dramatically improve the detection capability of the CMF as well as several examples of trained classifiers from the statistical learning literature. In some cases, the worst detection achieved by the SIDE filter significantly outperforms the CMF at its best. The trained classifiers work very well before enhancement. This may be a result of the approach taken by these techniques: they require some foreknowledge as to what a pixel with the target signal looks like in the current image. This availability of training data is not always an assumption that can be made. Despite the impressive performance of these techniques before enhancement, the B-SIDE filter significantly increases their detection power. This seems to underscore the reality that spatial information is being exploited and shows the value of these new SIDE filtering techniques.

The work that can be done in the future might include a better exploration of the choice of neighborhood size and shape. In this work, most of the experiments were done using a 3×3 square neighborhood. A smaller group of results using a 13-pixel neighborhood were presented in Chapter 4. More work could be done in exploring the value of these other neighborhood shapes, such as the 5-pixel neighborhood in fig. 2.3(c). These should be explored on objects that have characteristics that correspond to the supposed strengths of

these neighborhood shapes. There might be a simple horizontal or vertical extent that could be sought in a detection framework. Neighborhood shapes and sizes determine the complexity of the processing time. Because the LRs demand a sum be computed over half of the possible neighborhood configurations in the chosen neighborhood size, the number of terms increases by powers of 2. Quickly, this can become undesirable in size — a neighborhood with 13 pixels in it has $8,192$ configurations. As discussed in sec. 2.3.1, the computations can be cut in half so that $4,096$ terms are calculated for each pixel. For a 256×256 image, the number of terms becomes $268,435,456$.

More creative prior assignment can also be done. The training of the priors given some example training data set has been suggested. A further development on this idea would be to train two sets of priors: one for the neighborhoods which do have signal in the center pixel, and one for the neighborhoods which do not have signal in the center pixel. These would describe two different distributions. The distribution for signal present could be used in the numerator (the signal presence hypothesis test) and the distribution for signal absence could be used in the denominator. Other careful prior assignments could be made.

This work has focused on large billowy plumes that have significant contiguity. That is, these plumes have features that are much larger than the size of the neighborhood used to study them. Future efforts could include work on finding plumes or shapes that have features larger than a neighborhood in one dimension, but perhaps smaller features in the other dimension. Long, thin wisps of objects could be detected with enhanced detection capability with some careful work with the prior probability model and perhaps more sophisticated metrics to measure the "closeness" of a neighborhood to shapes that are desired to be emphasized.

Because this work has focused on large continuous plumes, a comparison to the effect of a 3×3 spatial averaging filter on the detection image (e.g., the CMF image) must be considered. The spatial averaging filter takes the average of the values in the pixels contained in the filter mask. This has the effect of a low-pass filter on the image content, so that high spatial frequencies are removed. In some cases, such as this one, where the plume is not a

high spatial frequency object, this low-pass filtering has a very positive effect on detection, as the curves show in fig. 5.1. The curve labelled "3×3" in those plots is the result of performing detection (generating ROC curves) on the CMF image that has been filtered by a 3×3 averaging filter. The other two curves are from SIDE using $\alpha = -100$ at two different neighborhood sizes. Note that the SIDE filter using a 3×3 neighborhood performs almost exactly as well as the spatial averaging filter. The SIDE filter using the larger 13-pixel neighborhood performs better than either at the release rates and values for SNR shown. This high value of α seems to cause the SIDE filter to behave much like a spatial averaging filter. This result will especially motive further exploration in the directions mentioned in this chapter: neighborhood shape and size, and prior assignment.

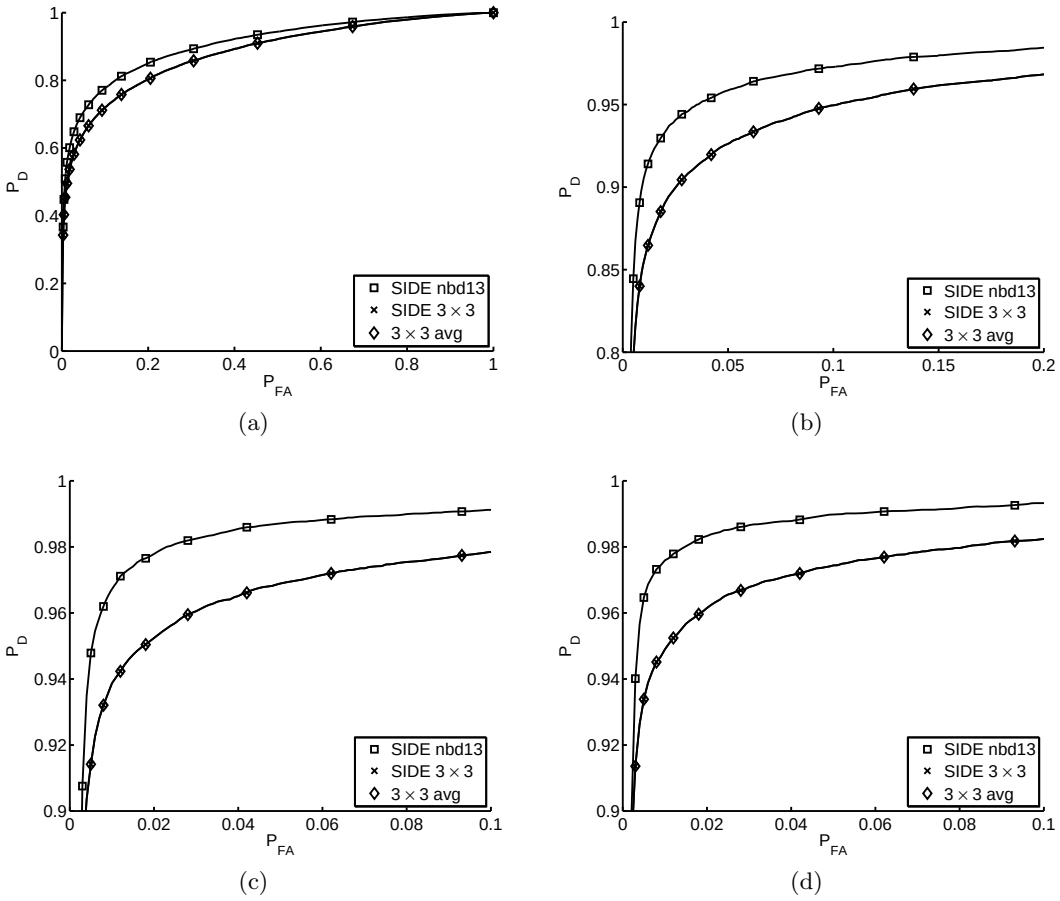


Fig. 5.1: SIDE filter using 3×3 neighborhood and 13-pixel neighborhood along with CMF + averaging filter. In (a), release rate = 100 ppm and SNR = 50 dB, (b) release rate = 100 ppm and SNR = 60 dB, (c) relase rate = 500 ppm and SNR = 50 dB, and (d) release rate = 500 ppm and SNR = 60 dB.

References

[1] *The DIRSIG User's Manual*, Rochester Institute of Technology, Oct. 2006 [Online]. Available: `www.dirsig.org/documentation`.

[2] D. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference*, vol. 2, pp. 1150–1157, 1999.

[3] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *Computer Vision and Pattern Recognition. Proceedings of the 2004 IEEE Computer Society Conference*, vol. 2, pp. 506–513, June-July 2004.

[4] W. M. Wells, III, "Statistical approaches to feature-based object recognition," *International Journal of Computer Vision*, vol. 21, no. 1-2, pp. 63–98, 1997.

[5] M. C. Burl, M. Weber, and P. Perona, "A probabilistic approach to object recognition using local photometry and global geometry," in *ECCV '98: Proceedings of the 5th European Conference on Computer Vision-Volume II*, pp. 628–641. London, UK: Springer-Verlag, 1998.

[6] G. Jaffre and A. Crouzil, "Non-rigid object localization from color model using mean shift," in *Image Processing. Proceedings of the 2003 International Conference*, vol. 3, pp. 317–320, Sept. 2003.

[7] B. R. Foy, "Overview of target detection algorithms," Los Alamos National Laboratory, Report to NNSA LA-UR-09-00593, June 2009.

[8] M. Bernhardt, J. P. Heather, and M. I. Smith, "New models for hyperspectral anomaly detection and un-mixing," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI. Proceedings SPIE*, vol. 5806, pp. 720–730. SPIE, 2005.

[9] P. Bajorski, "Analytical comparison of the matched filter and orthogonal subspace projection detectors in structured models for hyperspectral images," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII. Proceedings SPIE*, vol. 6233, no. 1. SPIE, 2006.

[10] P. Bajorski, E. J. Ientilucci, and J. R. Schott, "Comparison of basis-vector selection methods for target and background subspaces as applied to subpixel target detection," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X. Proceedings SPIE*, vol. 5425, pp. 97–108. SPIE, 2004.

[11] P. V. Villeneuve, H. A. Fry, J. P. Theiler, W. B. Clodius, B. W. Smith, and A. D. Stocker, "Improved matched-filter detection techniques," in *Imaging Spectrometry V. Proceedings SPIE*, vol. 3753, pp. 278–285. SPIE, 1999.

[12] J. Theiler and B. Foy, "Effect of signal contamination in matched-filter detection of the signal on a cluttered background," *Geoscience and Remote Sensing Letters, IEEE*, vol. 3, no. 1, pp. 98–102, Jan. 2006.

[13] Z. Wang, U. Sakoglu, S. Annamalai, N.-R. Weisse-Bernstein, P. Dowd, J. S. Tyo, M. M. Hayat, and S. Krishna, "Real-time implementation of matched filtering algorithms using adaptive focal-plane array technology," in *Imaging Spectrometry X. Proceedings SPIE*, vol. 5546, pp. 73–83. SPIE, 2004.

[14] T. Burr, B. Foy, H. Fry, and B. McVey, "Characterizing clutter in the context of detecting weak gaseous plumes in hyperspectral imagery," *Sensors*, vol. 6, no. 11, pp. 1587–1615, 2006.

[15] C. Funk, J. Theiler, D. Roberts, and C. Borel, "Clustering to improve matched filter detection of weak gas plumes in hyperspectral thermal imagery," *Geoscience and Remote Sensing, IEEE Transactions*, vol. 39, no. 7, pp. 1410–1420, July 2001.

[16] B. R. Foy and J. Theiler, "Scene analysis and detection in thermal infrared remote sensing using independent component analysis," in *Independent Component Analyses, Wavelets, Unsupervised Smart Sensors, and Neural Networks II. Proceedings SPIE*, vol. 5439, pp. 131–139. SPIE, 2004.

[17] G. Cao, C. A. Bouman, and J. Theiler, "Weak signal detection in hyperspectral imagery using sparse matrix transform (SMT) covariance estimation," in *Proceedings of WHISPERS, Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2009.

[18] N. B. Gallagher, D. M. Sheen, J. M. Shaver, B. M. Wise, and J. F. Shultz, "Estimation of trace vapor concentration pathlength in plumes for remote sensing applications from hyperspectral images," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX. Proceedings SPIE*, vol. 5093, pp. 184–194. SPIE, 2003.

[19] P. Bajorski, "Simplex projection methods for selection of endmembers in hyperspectral imagery," in *Geoscience and Remote Sensing Symposium. Proceedings of the 2004 IEEE International*, vol. 5, pp. 3207–3210, Sept. 2004.

[20] M. Dundar, J. Theiler, and S. Perkins, "Incorporating spatial contiguity into the design of a support vector machine classifier," in *Geoscience and Remote Sensing Symposium. Proceedings of the 2006 IEEE International*, pp. 364–367, Aug. 2006.

[21] M. Fauvel, J. Benediktsson, J. Chanussot, and J. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *Geoscience and Remote Sensing, IEEE Transactions*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[22] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A new method for target detection in hyperspectral imagery based on extended morphological profiles," in *Geoscience and Remote Sensing Symposium. Proceedings of the 2003 IEEE International*, vol. 6, pp. 3772–3774, July 2003.

[23] Y. Tarabalka, J. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *Geoscience and Remote Sensing, IEEE Transactions*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[24] M. Pesaresi and J. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *Geoscience and Remote Sensing, IEEE Transactions*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[25] P. Li and X. Xiao, "Evaluation of multiscale morphological segmentation of multispectral imagery for land cover classification," in *Geoscience and Remote Sensing Symposium. Proceedings of the 2004 IEEE International*, vol. 4, pp. 2676–2679, Sept. 2004.

[26] T. Moon, C. Grant, J. Gunther, and G. Williams, "A neighborhood model for detection in hyperspectral images," in *Signals, Systems and Computers, 2008 42nd Asilomar Conference*, pp. 1214–1218, Oct. 2008.

[27] T. Hastie, R. Tibshirani, and J. I. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, ser. Springer series in statistics. Berlin: Springer, 2001.

[28] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley-Interscience, 1973.

[29] I. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *Aerospace and Electronic Systems, IEEE Transactions*, vol. AES-10, no. 6, pp. 853–863, Nov. 1974.

[30] R. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.

[31] J. Theiler, B. R. Foy, and A. M. Fraser, "Beyond the adaptive matched filter: nonlinear detectors for weak signals in high-dimensional clutter," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII. Proceedings SPIE*, vol. 6565, no. 3. SPIE, 2007.

[32] J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*. New York: Cambridge University Press, 2006.

[33] Aviris free standard data products. Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA) [Online]. Available: `http://aviris.jpl.nasa.gov/html/aviris.freedata.html`.

[34] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 44, no. 2-3, pp. 127–143, 1993.

[35] J. Theiler, B. R. Foy, and A. M. Fraser, "Nonlinear signal contamination effects for gaseous plume detection in hyperspectral imagery," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII. Proceedings SPIE*, vol. 6233, no. 1U. SPIE, 2006.

[36] J. A. Hackwell, D. W. Warren, R. P. Bongiovi, S. J. Hansel, T. L. Hayhurst, D. J. Mabry, M. G. Sivjee, and J. W. Skinner, "LWIR/MWIR imaging hyperspectral sensor for airborne and ground-based remote sensing," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 2819, pp. 102–107, Nov. 1996.

[37] E. R. Keim, L. E. Kirkland, J. A. Hackwell, and K. C. Herr, "Terrestrial Airborne Hyperspectral Remote Sensing (SEBASS): Applications to remote sensing of Mars," in *Proceedings of the 2001 Lunar and Planetary Science Conference*, vol. 32, no. 2162, Mar. 2001.

[38] E. M. O'Donnell, D. W. Messinger, C. Salvaggio, and J. R. Schott, "Identification and detection of gaseous effluents from hyperspectral imagery using invariant algorithms," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X. Proceedings SPIE*, vol. 5425, pp. 573–582. SPIE, 2004.

[39] J. Theiler, B. R. Foy, and A. M. Fraser, "Characterizing non-gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery," vol. 5806, no. 1, pp. 182–193. SPIE, 2005.

[40] M. K. Griffin, R. N. Czerwinski, C. A. Upham, E. C. Wack, and H. hua K. Burke, "A procedure for embedding effluent plumes into LWIR imagery," vol. 5806, no. 1, pp. 78–87. SPIE, 2005.

[41] S. D. Kuo, J. R. Schott, and C. Y. Chang, "Synthetic image generation of chemical plumes for hyperspectral applications," *Optical Engineering*, vol. 39, no. 4, pp. 1047–1056, 2000.

[42] A. Berk and G. Anderson, "Impact of MODTRAN^{TM}5.1 on atmospheric compensation," in *Geoscience and Remote Sensing Symposium, 2008. IEEE International*, vol. 3, pp. 127–129, July 2008.

[43] A. K. Blackadar, *Turbulence and Diffusion in the Atmosphere*. Berlin: Springer, 1997.

[44] J. Nascimento and J. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

Appendices

Appendix A

Complete SIDE Filter Results

Whereas only a portion of the experimental results for the SIDE filter are shown in Chapter 4, the complete results are shown here. These results are grouped in plots showing a pair of curves (one for the CMF, one for the SIDE filter) for each release rate given a set of imaging circumstances based on atmosphere, sensor elevation, SNR, and the value of α. Each page shows four plots: one for each value of α tested.

A pair of tables are also included in this appendix, showing all these same results in the form of areas under the ROC curve. There are two tables, one for each atmosphere used in the experiments. These tables are found in sec. A.2.

A.1 ROC Curves

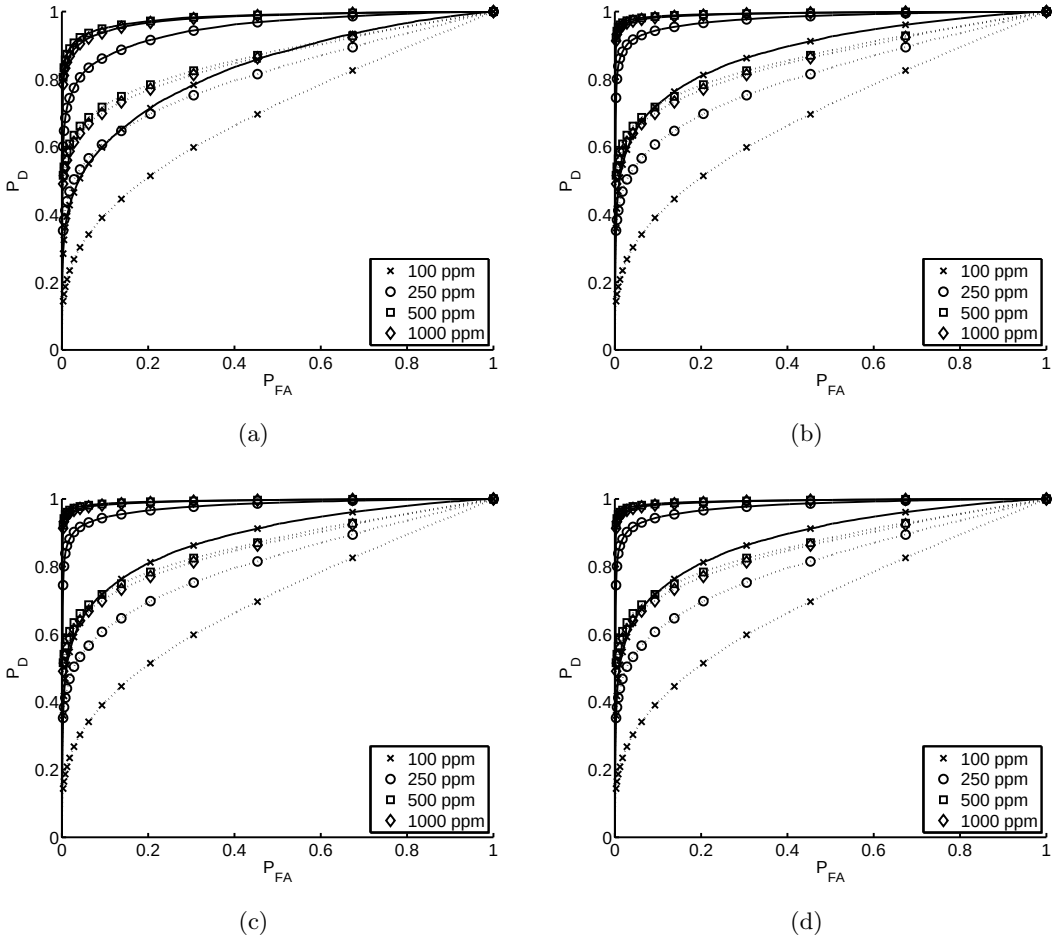


Fig. A.1: SIDE filter vs. CMF at mid-lat summer, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

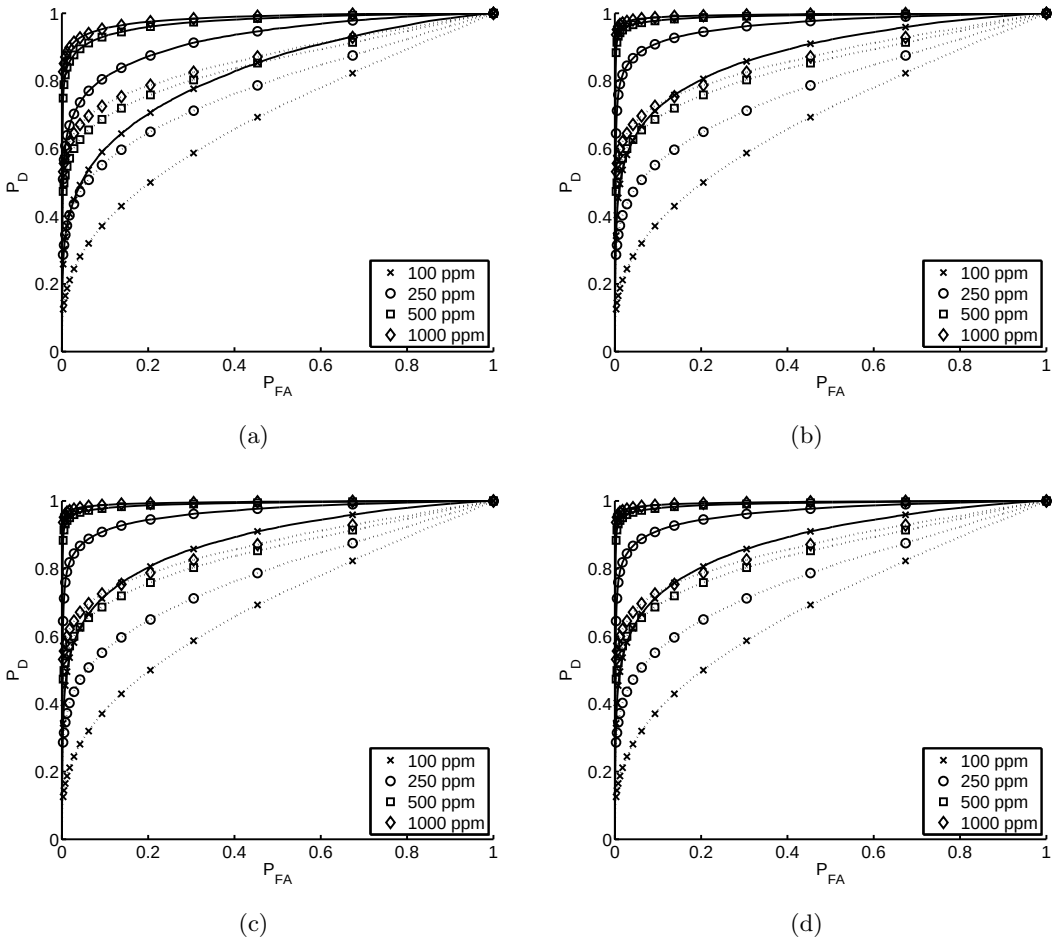


Fig. A.2: SIDE filter vs. CMF at mid-lat summer, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

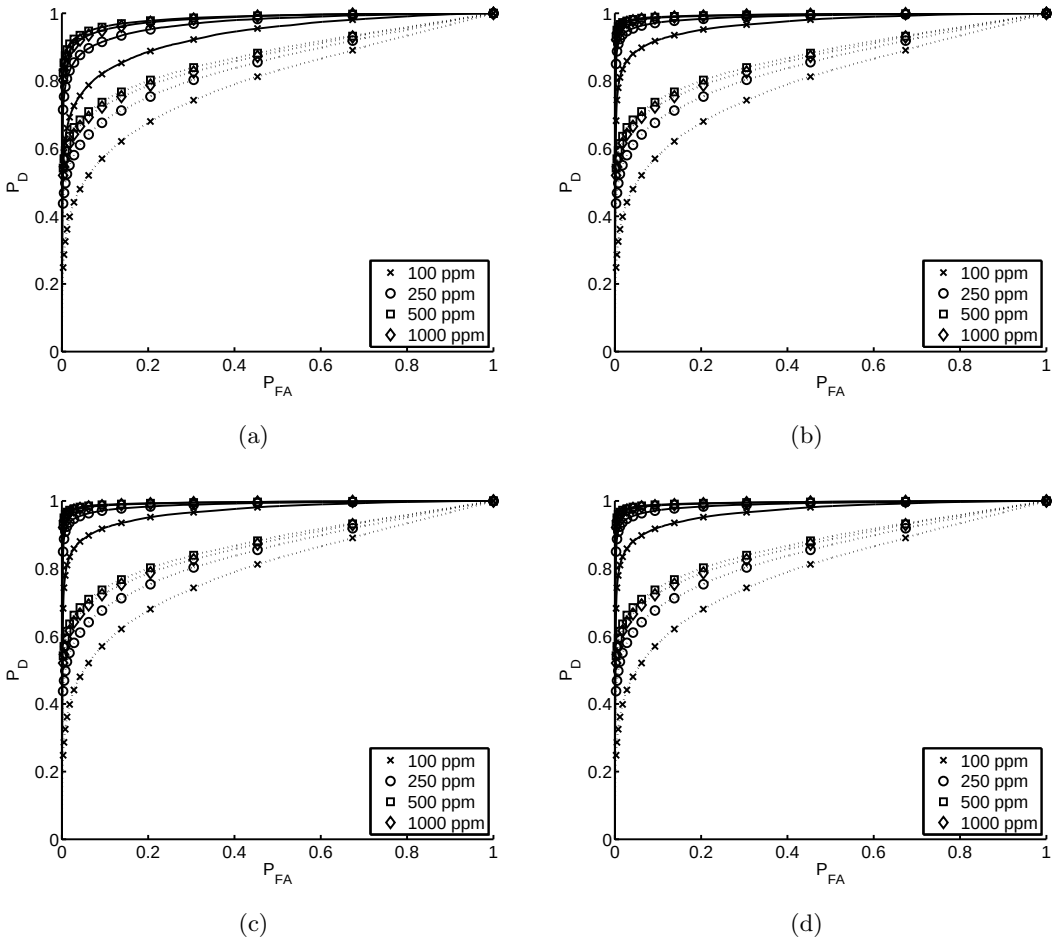


Fig. A.3: SIDE filter vs. CMF at mid-lat winter, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

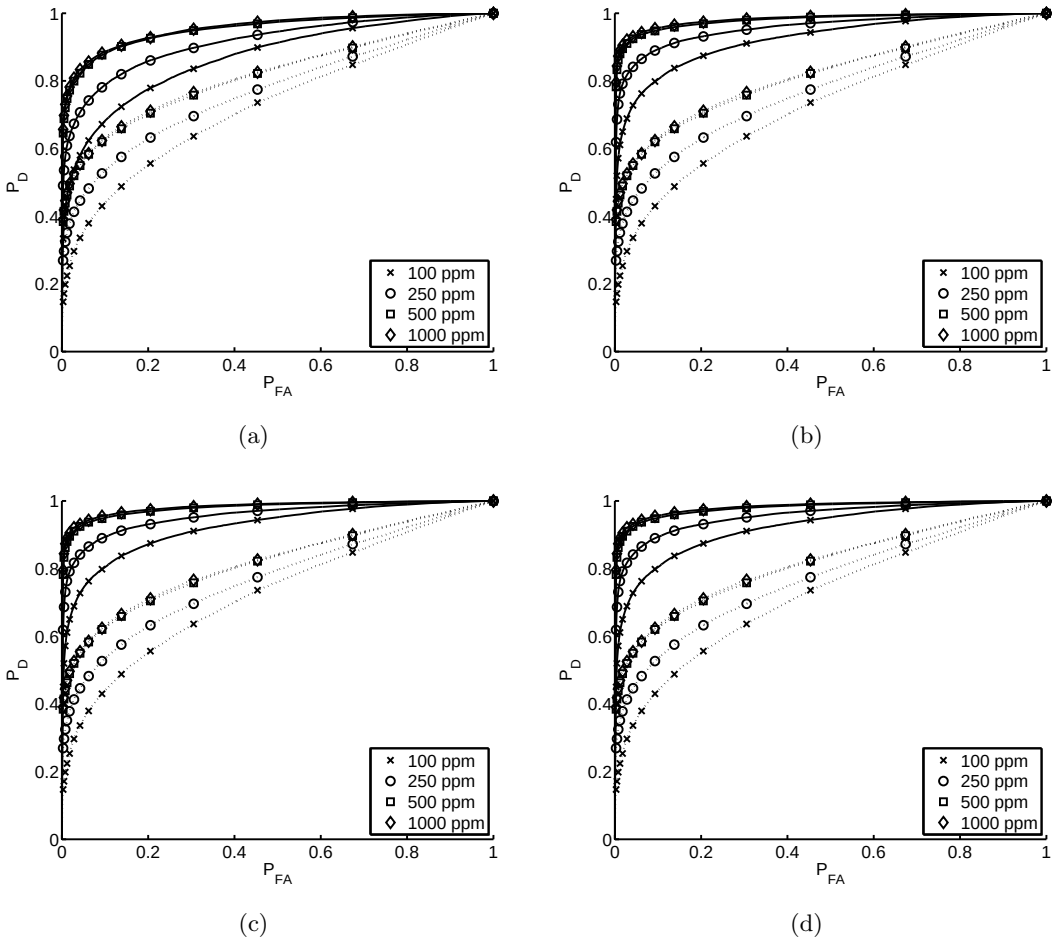


Fig. A.4: SIDE filter vs. CMF at mid-lat winter, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

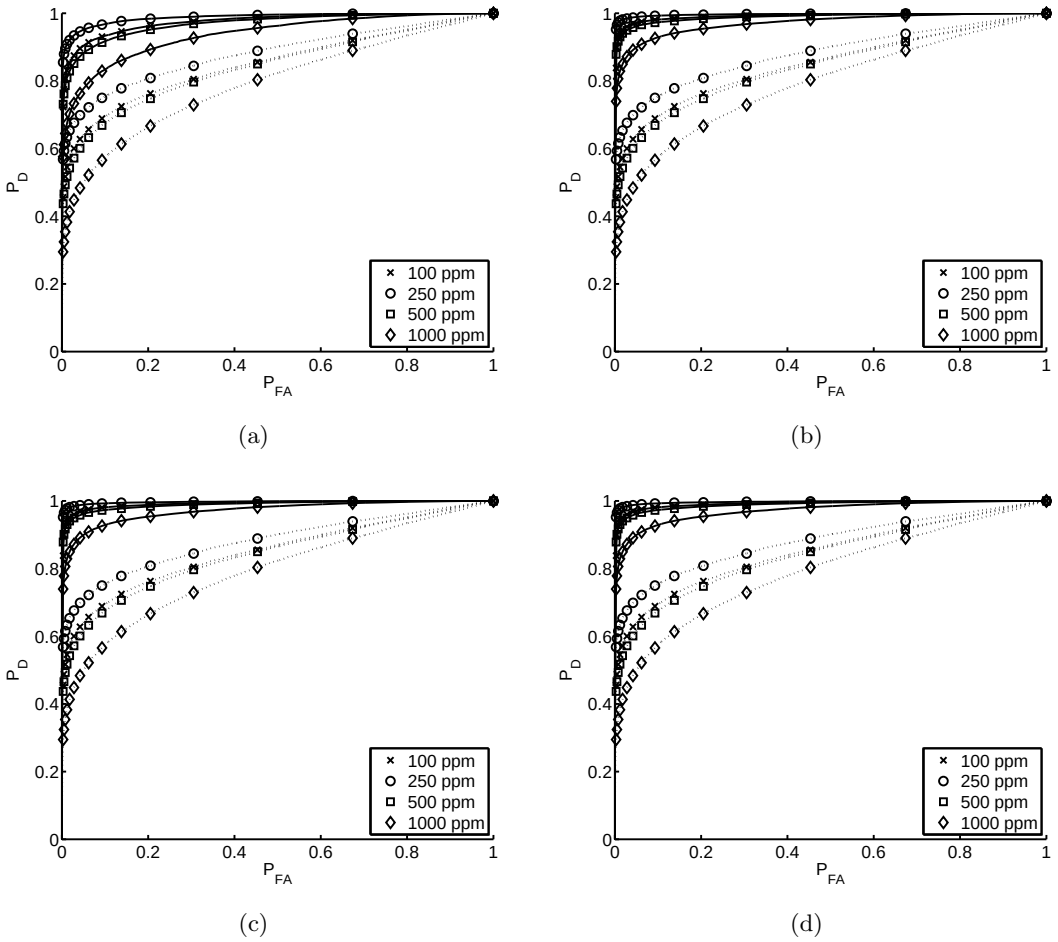


Fig. A.5: SIDE filter vs. CMF at mid-lat summer, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

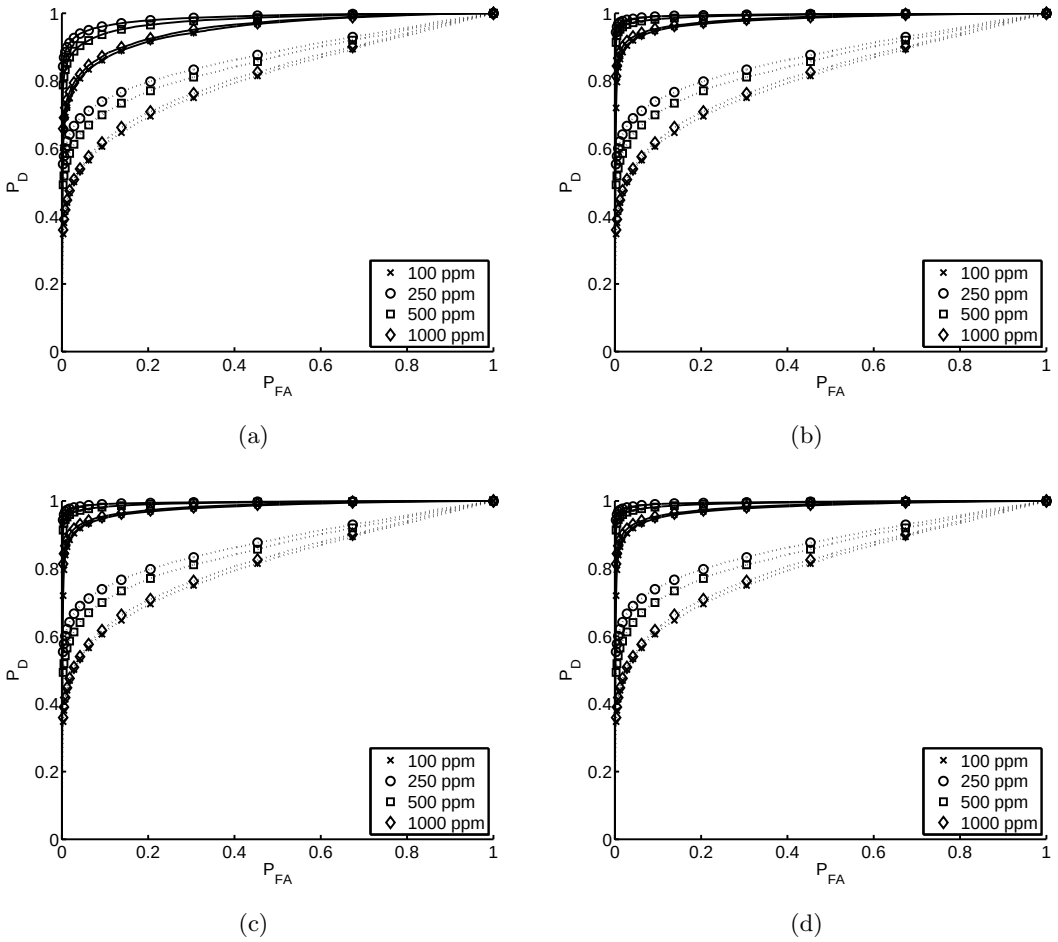


Fig. A.6: SIDE filter vs. CMF at mid-lat summer, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

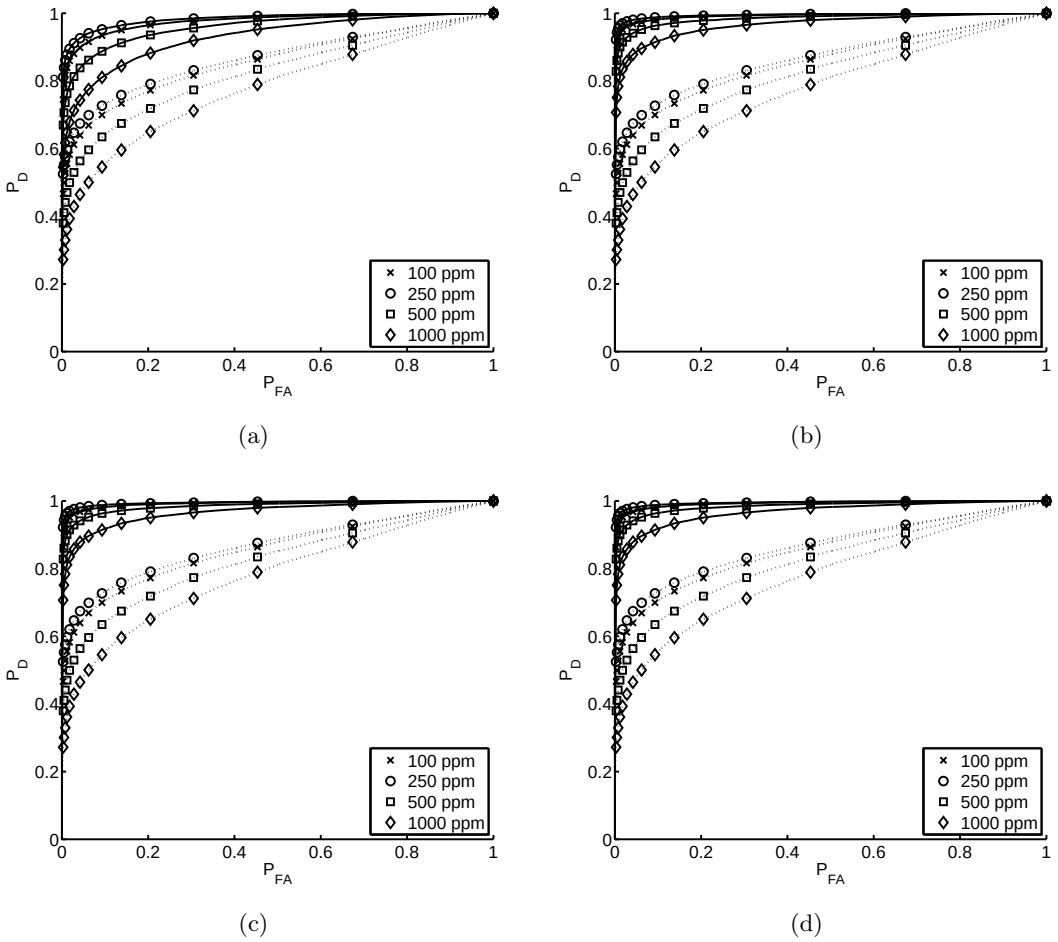


Fig. A.7: SIDE filter vs. CMF at mid-lat winter, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

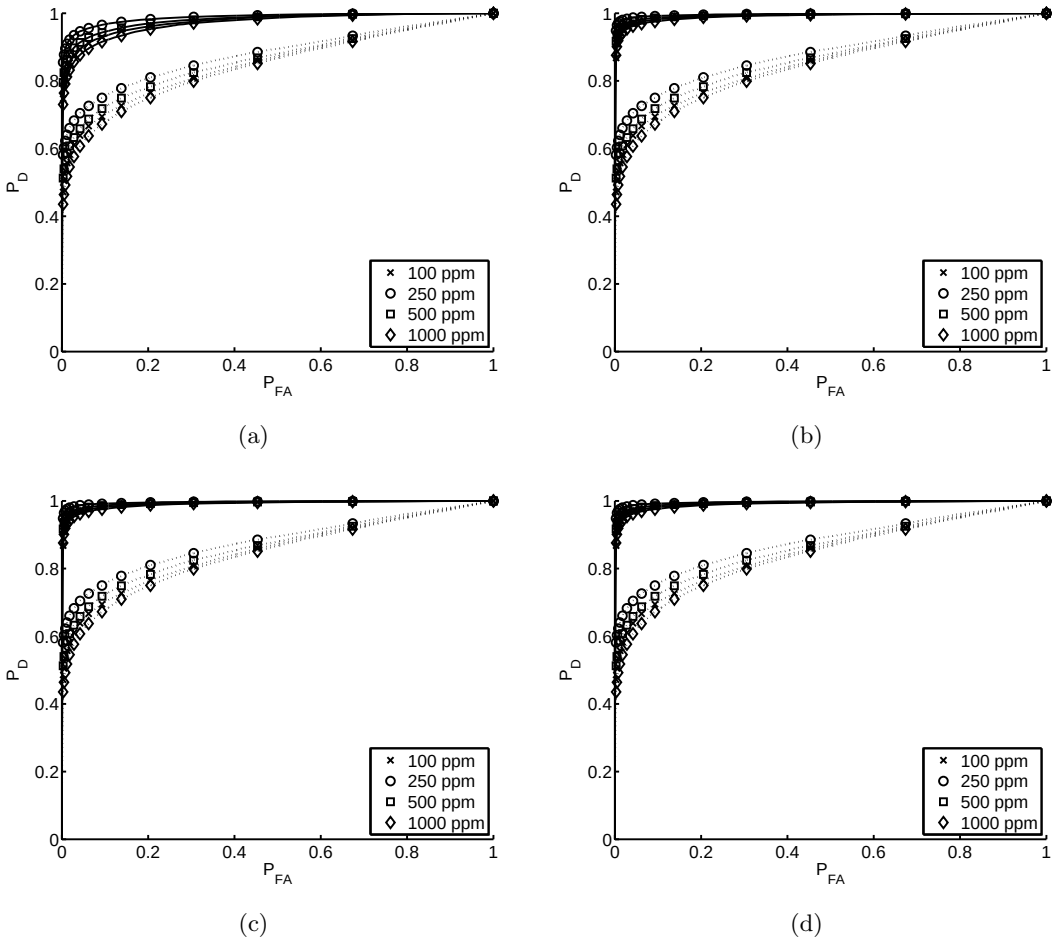


Fig. A.8: SIDE filter vs. CMF at mid-lat winter, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

A.2 Tables

Table A.1: Area under ROC curve for CMF and SIDE at mid-lat summer.

				A_S, $\alpha =$			
SNR (dB)	elevation (m)	rate (ppm)	A_C	1	10	10^2	10^3
50	5k	100	0.69677	0.83227	0.88667	0.88667	0.88667
		250	0.80898	0.94804	0.97747	0.97747	0.97747
		500	0.86434	0.98102	0.99358	0.99358	0.99358
		1000	0.85596	0.97762	0.99258	0.99258	0.99258
	10k	100	0.68938	0.82719	0.88336	0.88336	0.88336
		250	0.77934	0.92452	0.96366	0.96366	0.96366
		500	0.8467	0.9734	0.99042	0.99041	0.99041
		1000	0.86668	0.98306	0.99467	0.99467	0.99467
60	5k	100	0.85004	0.97506	0.9914	0.9914	0.9914
		250	0.88115	0.98798	0.99653	0.99653	0.99653
		500	0.84162	0.96896	0.98885	0.98885	0.98885
		1000	0.79257	0.936	0.97086	0.97086	0.97086
	10k	100	0.80803	0.9487	0.97816	0.97816	0.97816
		250	0.87222	0.98541	0.99526	0.99526	0.99526
		500	0.85376	0.97679	0.99255	0.99255	0.99255
		1000	0.81695	0.95417	0.98207	0.98207	0.98207

Table A.2: Area under ROC curve for CMF and SIDE at mid-lat winter.

				A_S, $\alpha =$			
SNR (dB)	elevation (m)	rate (ppm)	A_C	1	10	10^2	10^3
50	5k	100	0.79717	0.93213	0.96831	0.96831	0.96831
		250	0.8462	0.96895	0.98813	0.98813	0.98813
		500	0.87473	0.98467	0.99449	0.99449	0.99449
		1000	0.86662	0.98173	0.99377	0.99377	0.99377
	10k	100	0.7242	0.86912	0.92205	0.92205	0.92205
		250	0.76873	0.91452	0.9558	0.9558	0.9558
		500	0.81484	0.95278	0.97948	0.97948	0.97948
		1000	0.8186	0.95653	0.9824	0.9824	0.9824
60	5k	100	0.85696	0.97616	0.99183	0.99183	0.99183
		250	0.86872	0.98317	0.99461	0.99461	0.99461
		500	0.8244	0.95963	0.98455	0.98455	0.98455
		1000	0.78036	0.92897	0.96681	0.96681	0.96681
	10k	100	0.85245	0.97484	0.99194	0.99194	0.99194
		250	0.87959	0.98751	0.99632	0.99632	0.99632
		500	0.86284	0.97976	0.99374	0.99374	0.99374
		1000	0.84348	0.97003	0.9903	0.9903	0.9903

Appendix B

B–SIDE Filter Results

Whereas only a portion of the experimental results for the B–SIDE filter are shown in Chapter 4, the complete results are shown here. These results are grouped in plots showing a pair of curves (one for the baseline technique, one for the B–SIDE + baseline results) for each release rate given a set of imaging circumstances based on atmosphere, sensor elevation, SNR, and the value of α. Each page shows four plots: one for each value of α tested.

A pair of tables for each baseline technique are also included in this appendix, showing all these same results in the form of areas under the ROC curve. There are two tables for each baseline technique, one for each atmosphere used in the experiments. These tables are found in the respective section for each baseline technique.

B.1 CMF

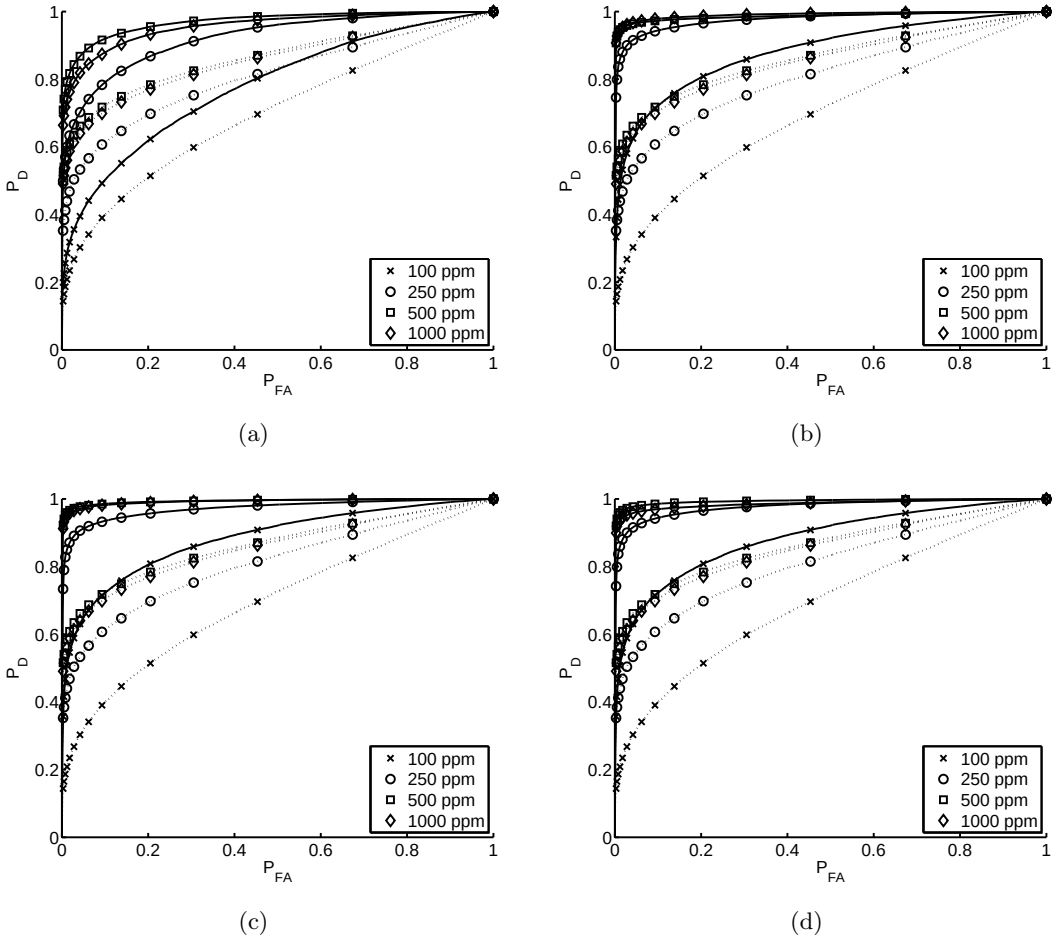


Fig. B.1: B–SIDE filter and CMF at mid-lat summer, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

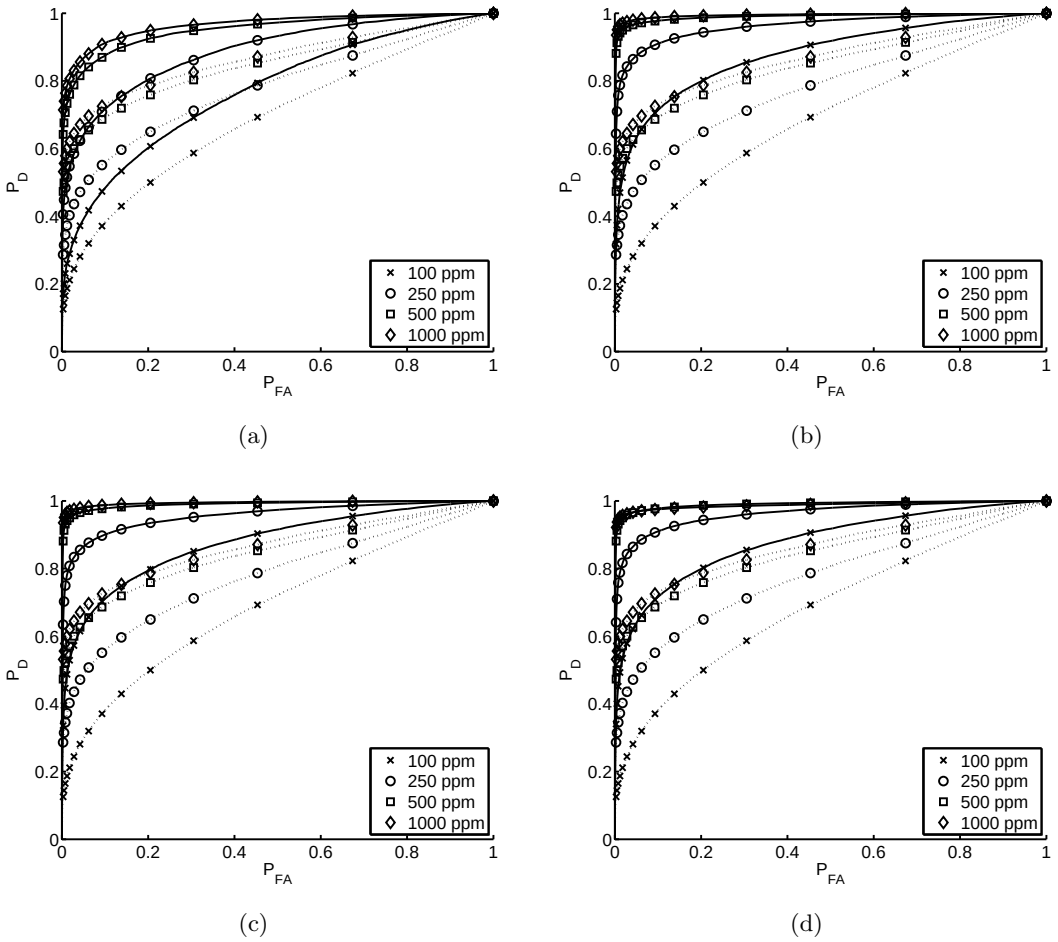


Fig. B.2: B–SIDE filter and CMF at mid-lat summer, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

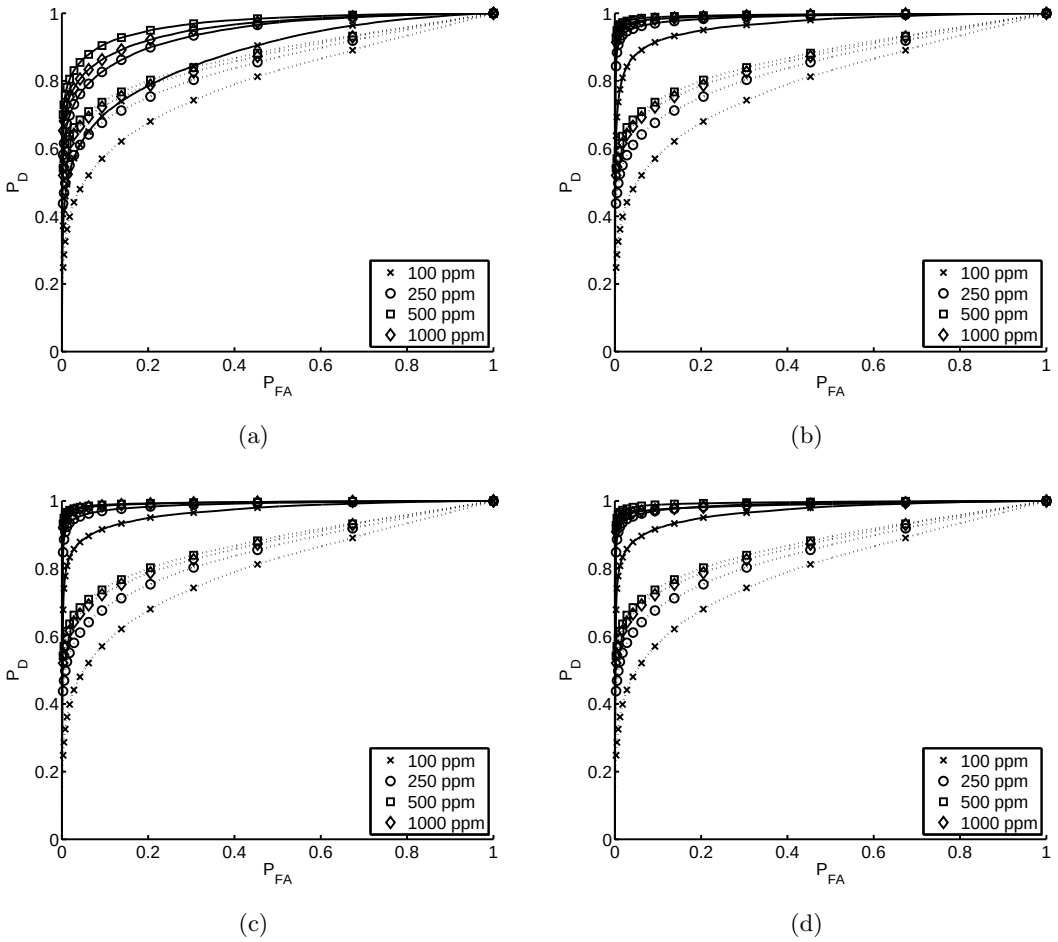


Fig. B.3: B–SIDE filter and CMF at mid-lat winter, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

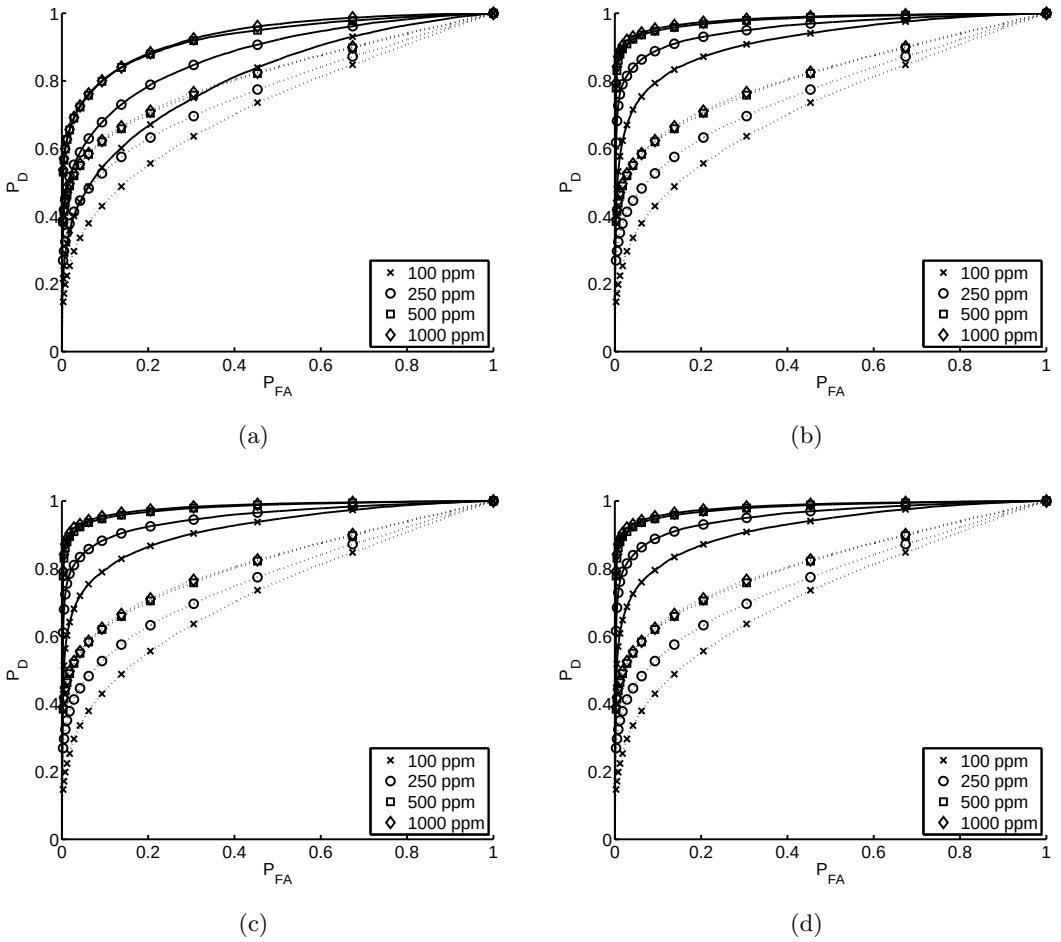


Fig. B.4: B–SIDE filter and CMF at mid-lat winter, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

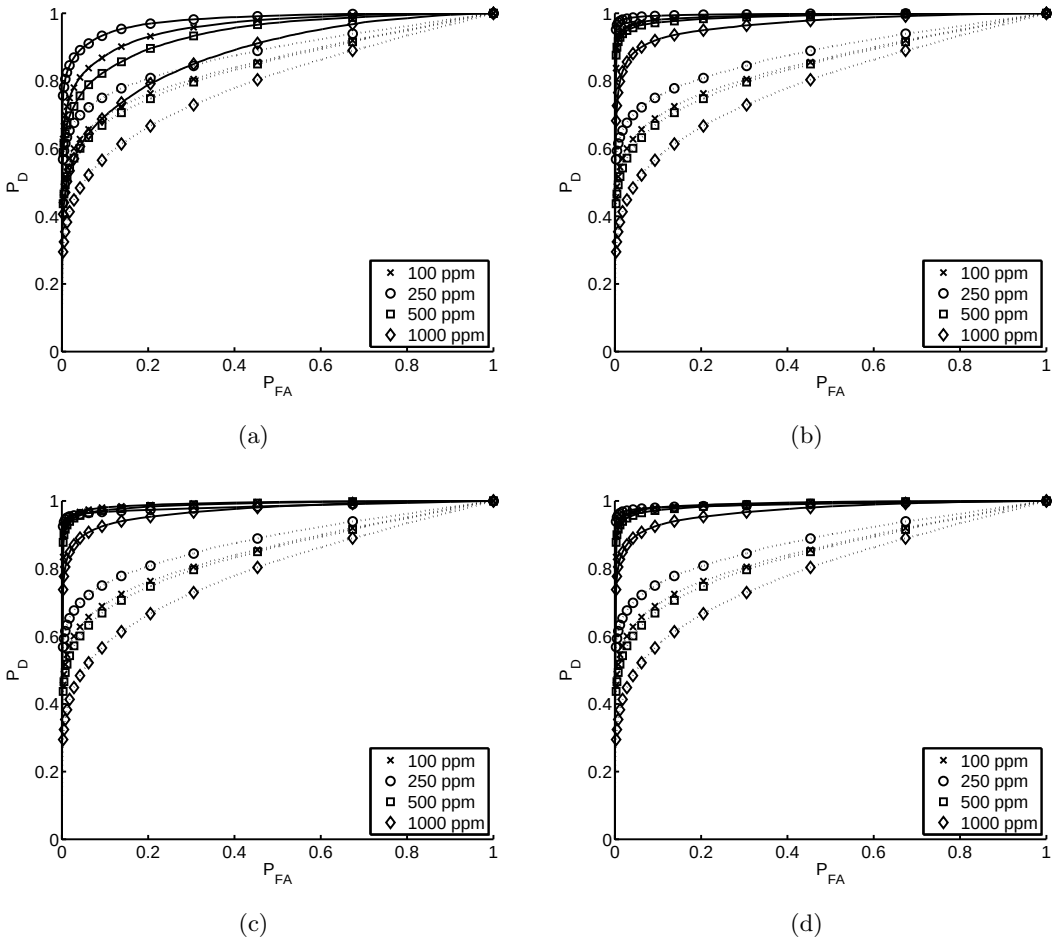


Fig. B.5: B–SIDE filter and CMF at mid-lat summer, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

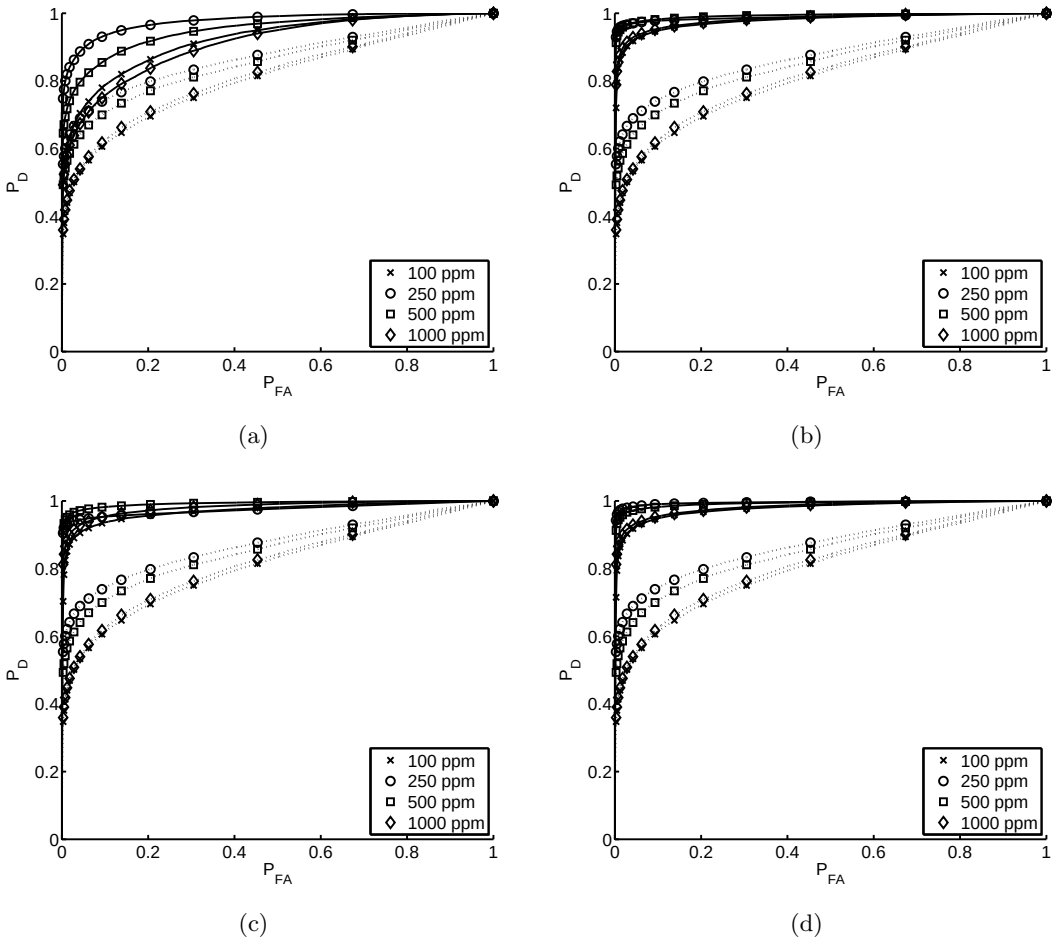


Fig. B.6: B–SIDE filter and CMF at mid-lat summer, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

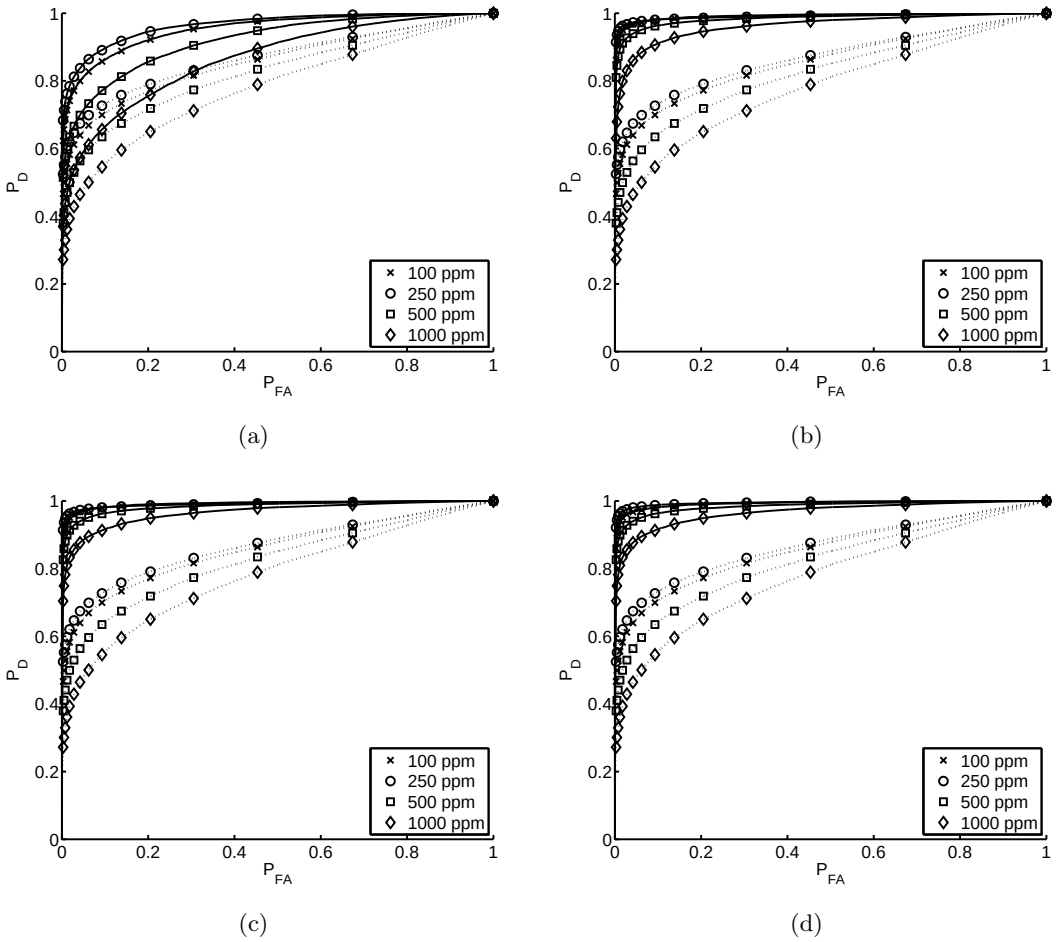


Fig. B.7: B–SIDE filter and CMF at mid-lat winter, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

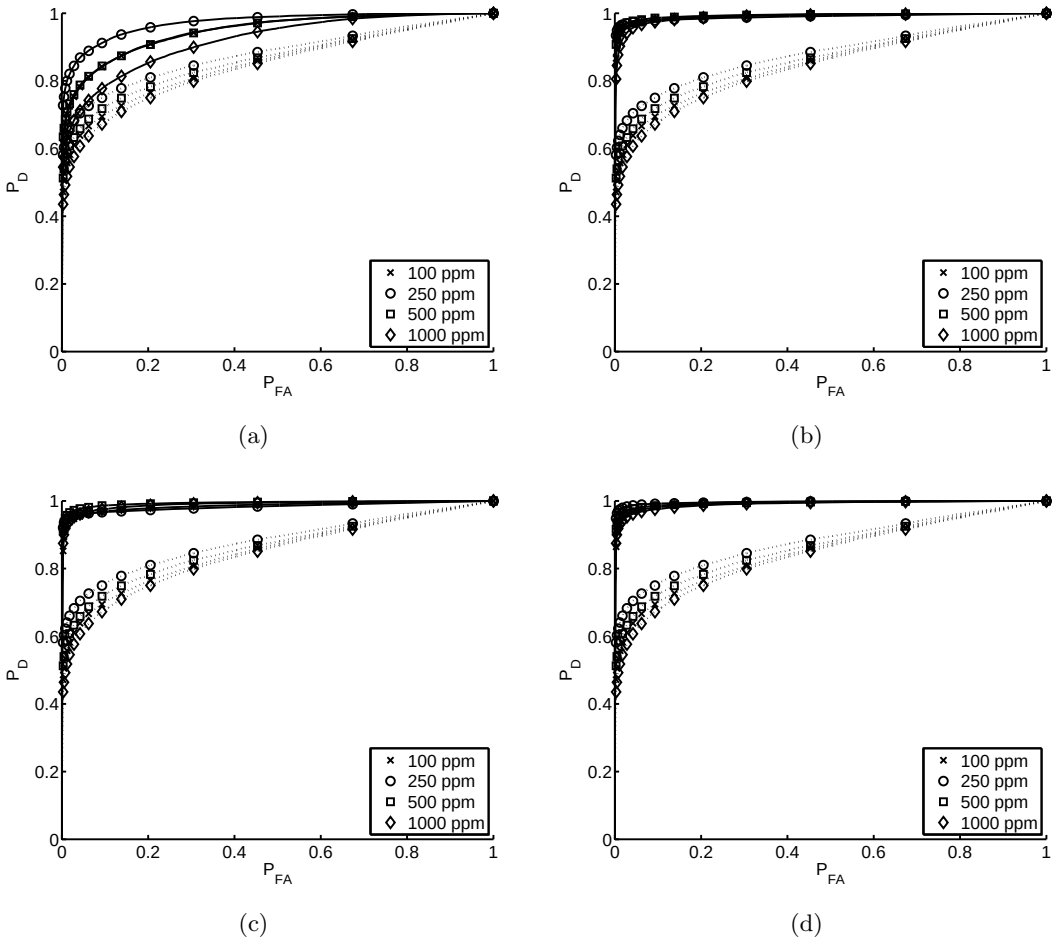


Fig. B.8: B–SIDE filter and CMF at mid-lat winter, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

Table B.1: Area under ROC curve for CMF and B–SIDE + CMF at mid-lat summer.

				A_{B_C}, $\alpha =$			
SNR (dB)	elevation (m)	rate (ppm)	A_C	1	10	10^2	10^3
50	5k	100	0.69677	0.78082	0.88305	0.88369	0.88369
		250	0.80898	0.92227	0.97689	0.97171	0.97678
		500	0.86434	0.97014	0.98647	0.99336	0.99336
		1000	0.85596	0.95587	0.99077	0.99233	0.98547
	10k	100	0.68938	0.77127	0.87927	0.87769	0.88032
		250	0.77934	0.88883	0.96267	0.95704	0.9626
		500	0.8467	0.9512	0.99012	0.99009	0.99009
		1000	0.86668	0.96561	0.99448	0.99448	0.98755
60	5k	100	0.85004	0.9569	0.99127	0.99112	0.99112
		250	0.88115	0.97785	0.99644	0.98258	0.98951
		500	0.84162	0.93875	0.98854	0.98851	0.98851
		1000	0.79257	0.8814	0.96712	0.97005	0.97005
	10k	100	0.80803	0.92095	0.97763	0.97074	0.97749
		250	0.87222	0.97597	0.98818	0.97464	0.9951
		500	0.85376	0.94845	0.99234	0.99232	0.99232
		1000	0.81695	0.90933	0.98134	0.98155	0.98155

Table B.2: Area under ROC curve for CMF and B–SIDE + CMF at mid-lat winter.

				A_{B_C}, $\alpha =$			
SNR (dB)	elevation (m)	rate (ppm)	A_C	1	10	10^2	10^3
50	5k	100	0.79717	0.87877	0.96624	0.96742	0.96742
		250	0.8462	0.93971	0.98784	0.98777	0.98777
		500	0.87473	0.96705	0.99441	0.99432	0.99432
		1000	0.86662	0.952	0.99209	0.99358	0.98719
	10k	100	0.7242	0.81064	0.91875	0.91706	0.91998
		250	0.76873	0.87664	0.95462	0.95106	0.95458
		500	0.81484	0.92535	0.97893	0.97885	0.97885
		1000	0.8186	0.93135	0.98195	0.98185	0.98185
60	5k	100	0.85696	0.95269	0.99167	0.99157	0.99157
		250	0.86872	0.96476	0.99053	0.99066	0.99444
		500	0.8244	0.91912	0.98398	0.9841	0.9841
		1000	0.78036	0.86594	0.9624	0.9659	0.9659
	10k	100	0.85245	0.9467	0.99179	0.98482	0.99169
		250	0.87959	0.97212	0.98931	0.98238	0.99621
		500	0.86284	0.94646	0.99354	0.99356	0.99356
		1000	0.84348	0.91922	0.98898	0.99003	0.99003

B.2 LDA

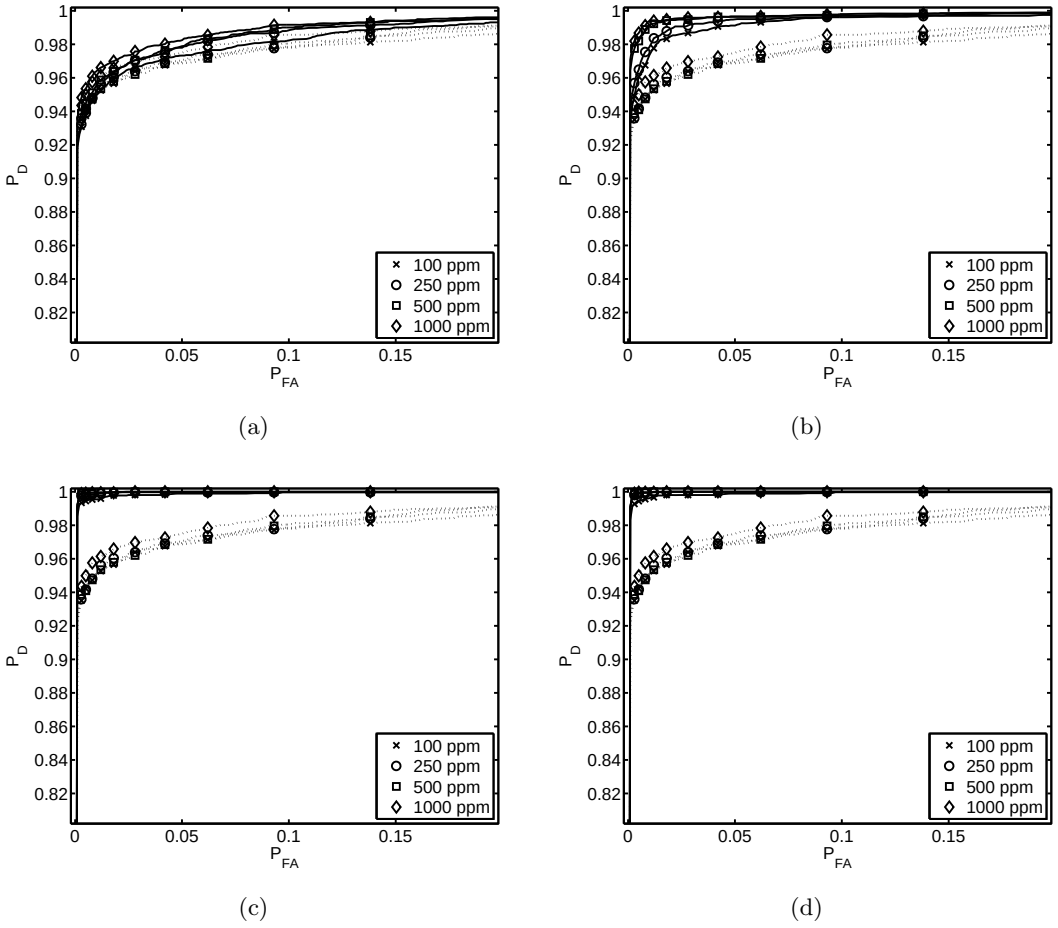


Fig. B.9: B–SIDE filter and LDA at mid-lat summer, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

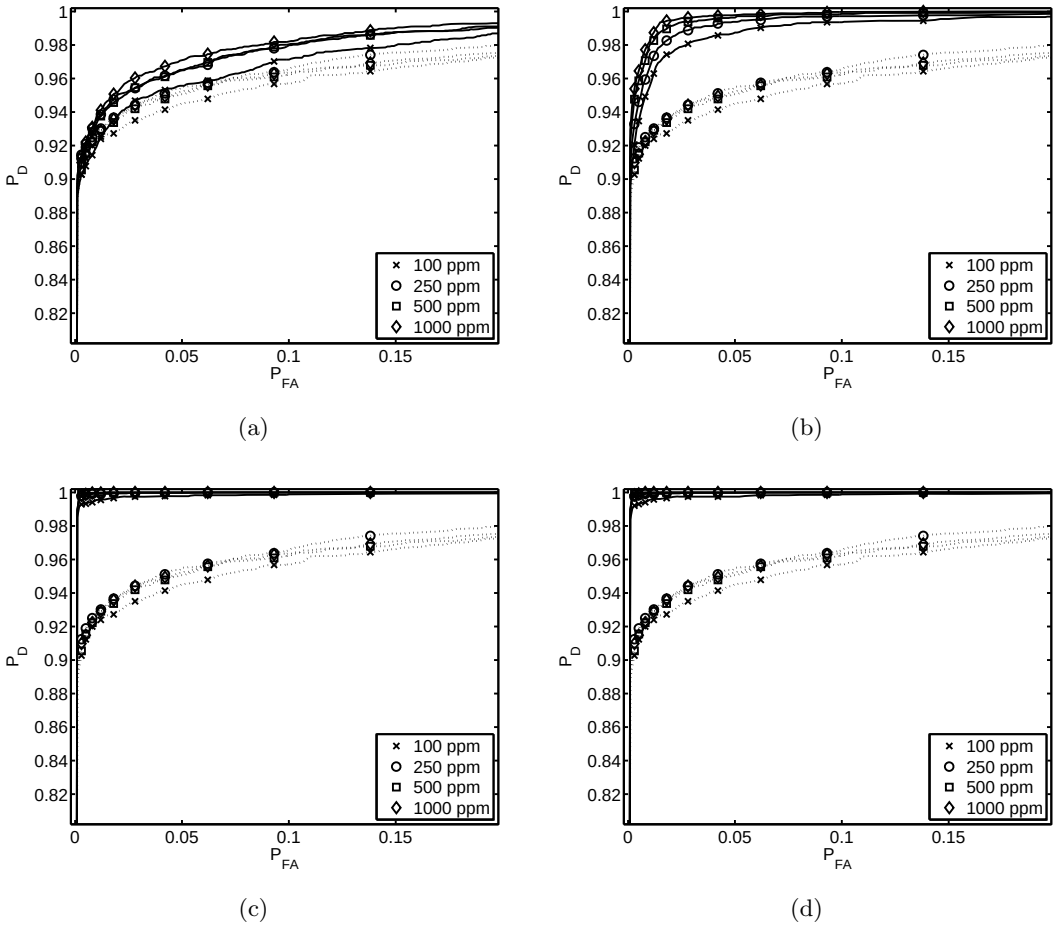


Fig. B.10: B–SIDE filter and LDA at mid-lat summer, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

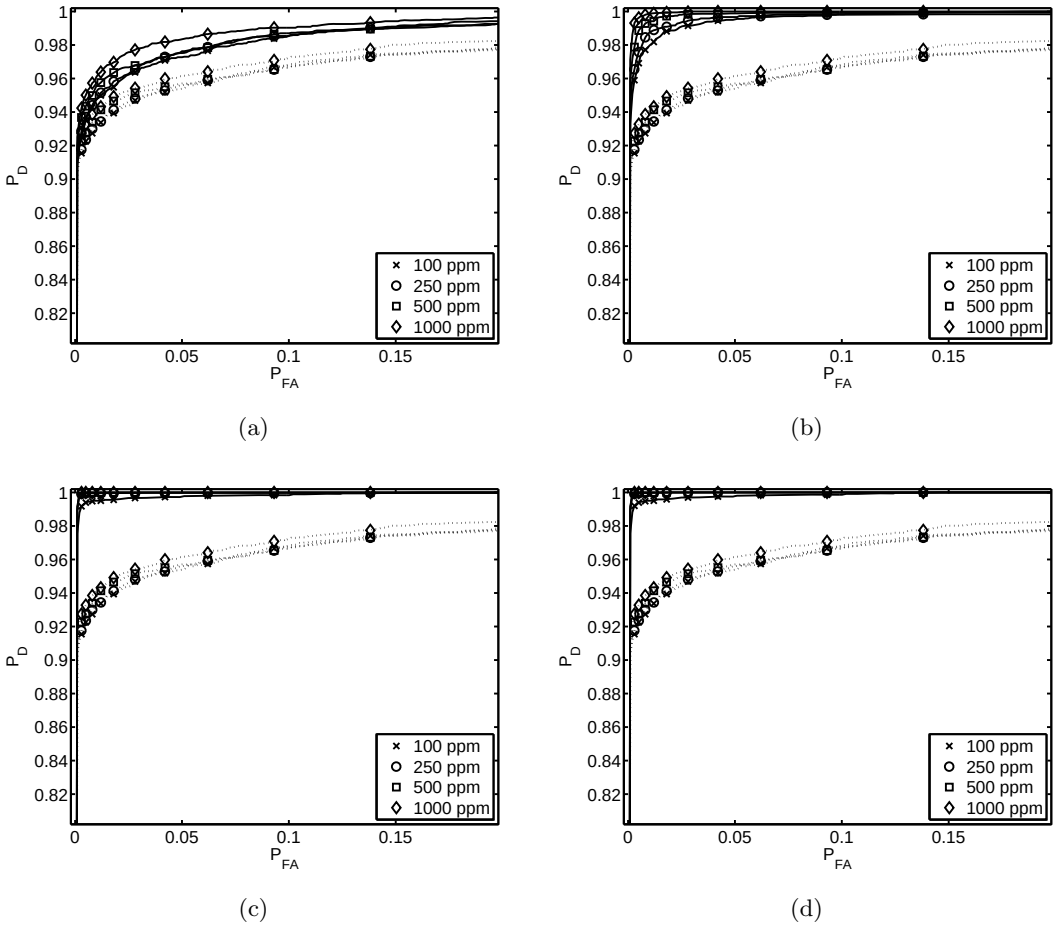


Fig. B.11: B–SIDE filter and LDA at mid-lat winter, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

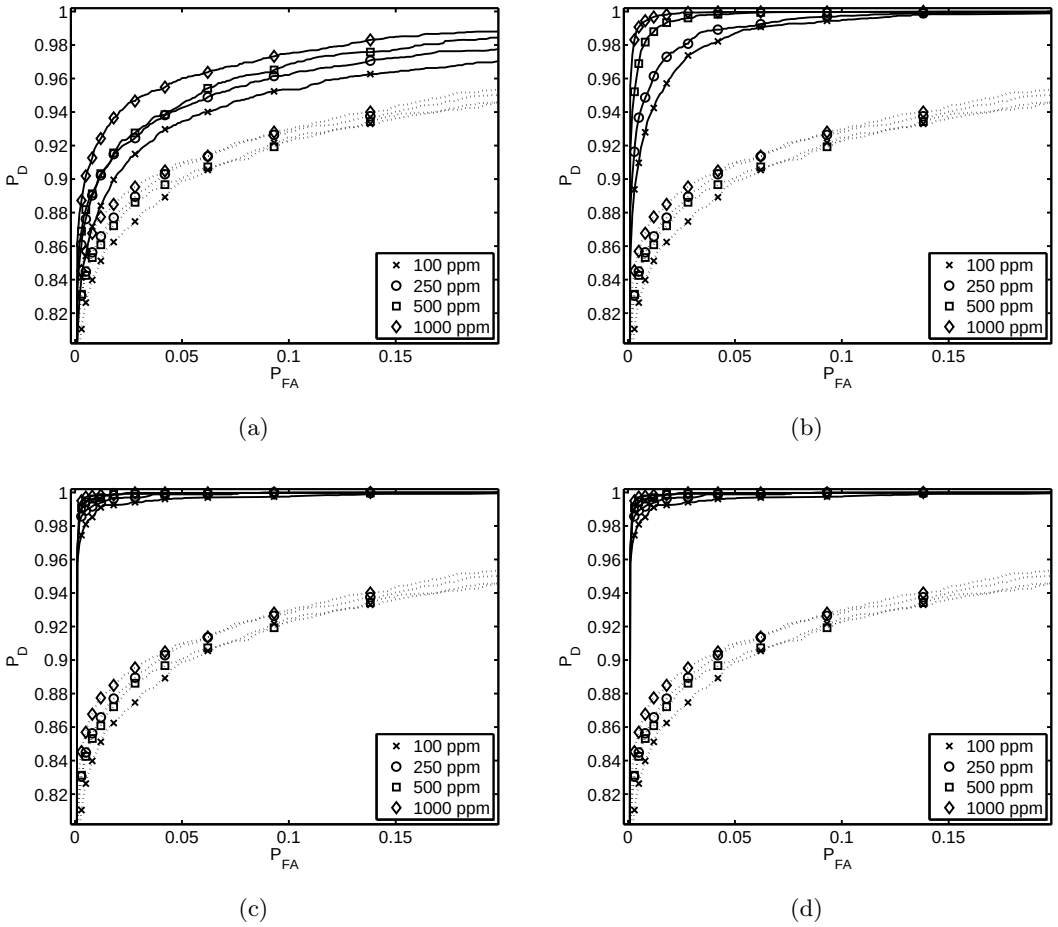


Fig. B.12: B–SIDE filter and LDA at mid-lat winter, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

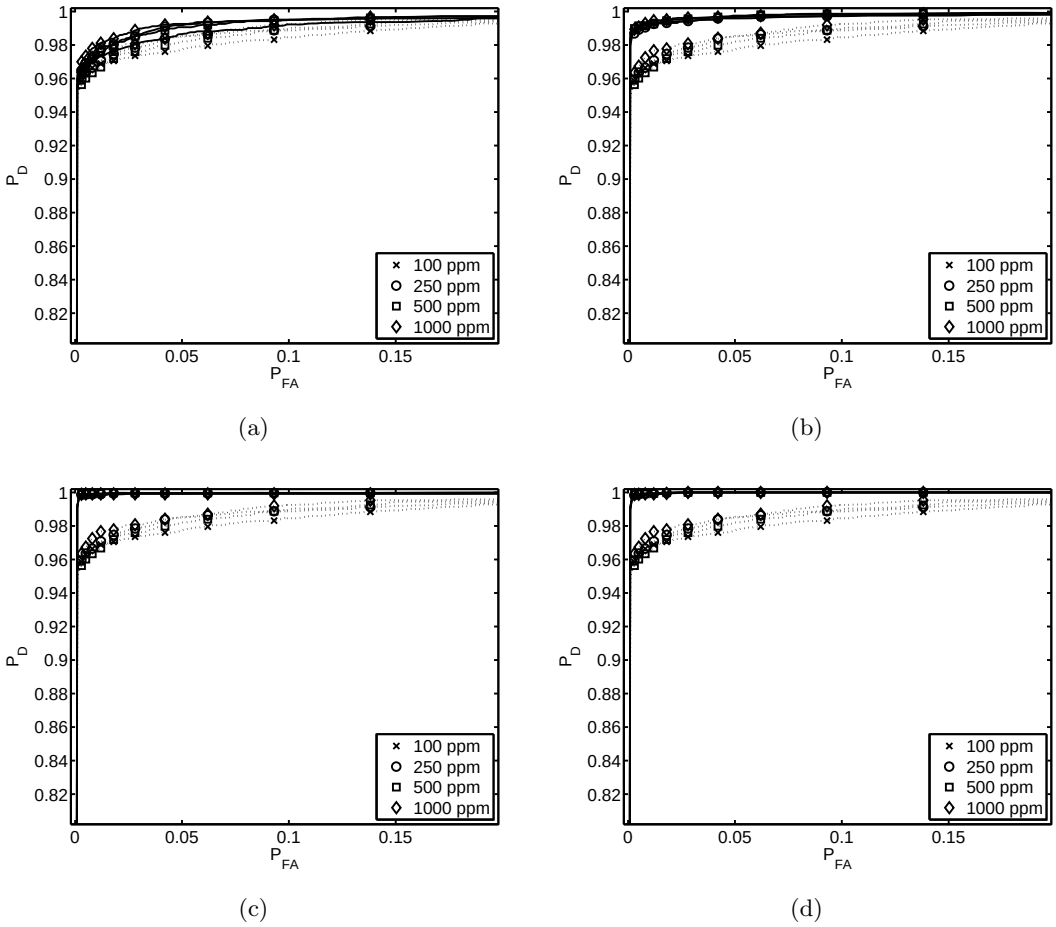


Fig. B.13: B–SIDE filter and LDA at mid-lat summer, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

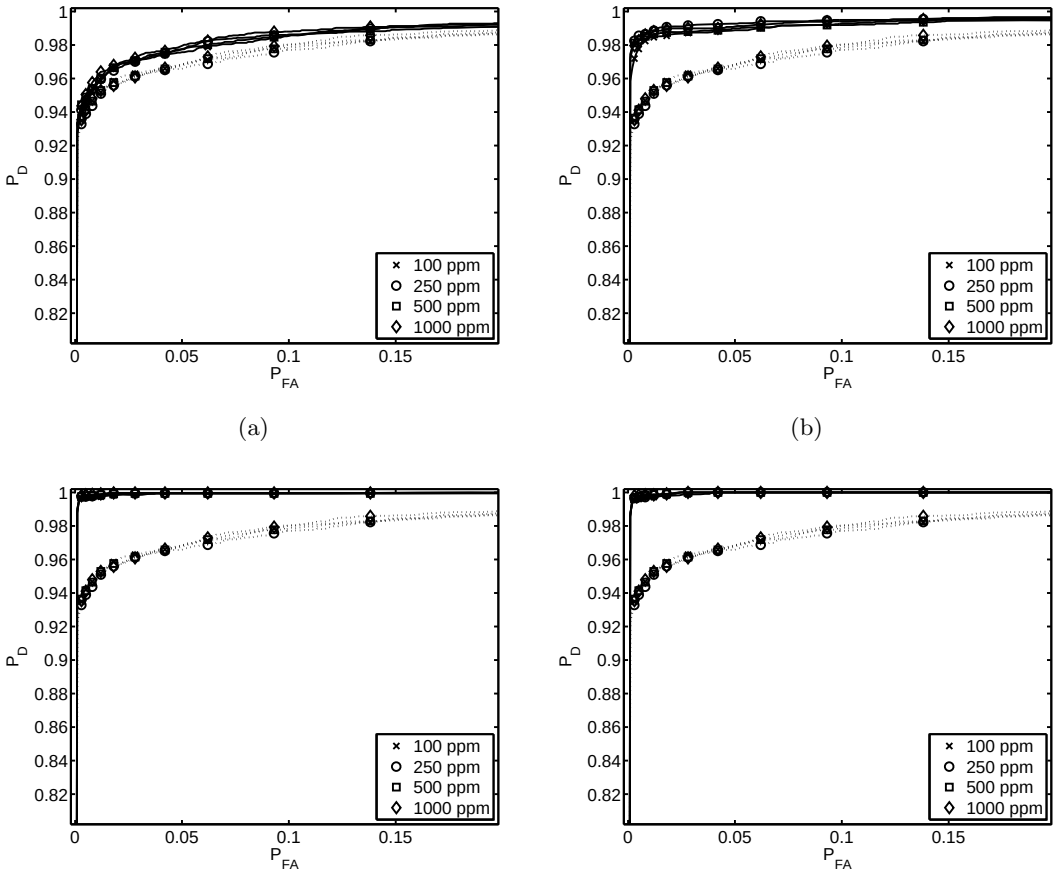


Fig. B.14: B–SIDE filter and LDA at mid-lat summer, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

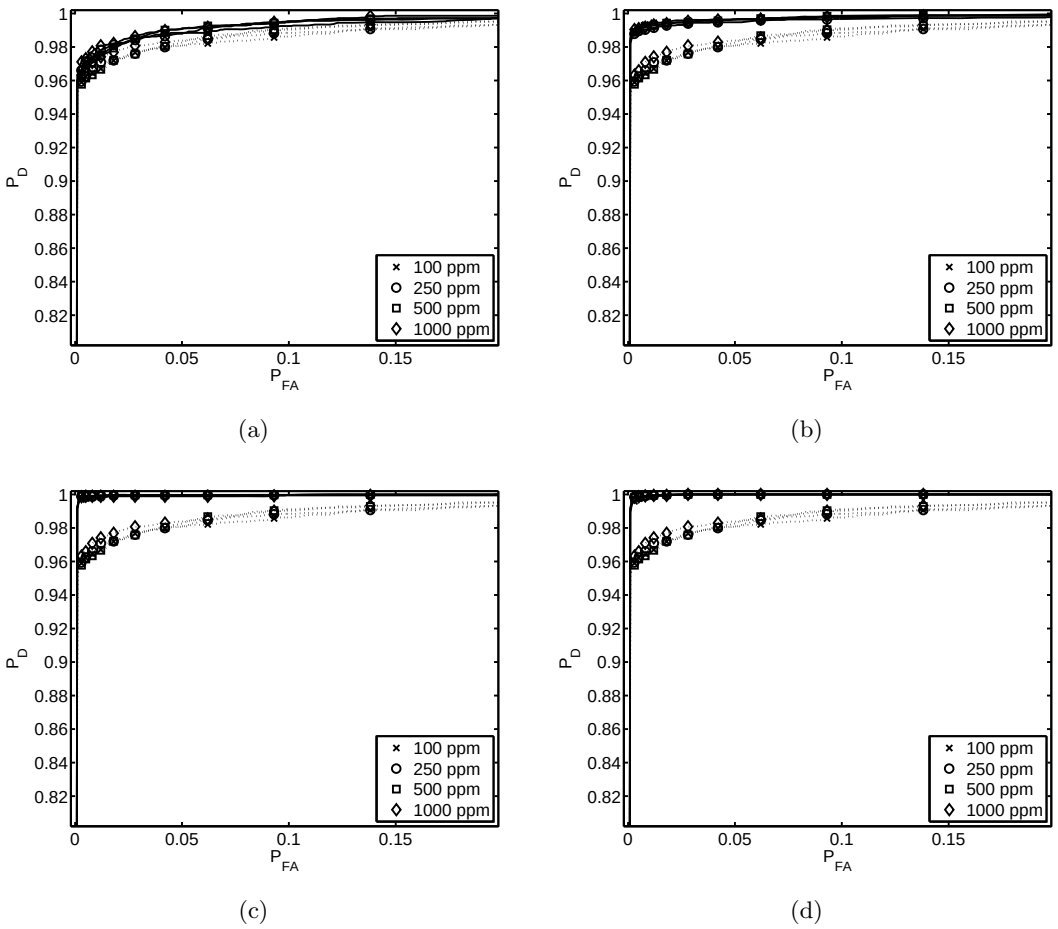


Fig. B.15: B–SIDE filter and LDA at mid-lat winter, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

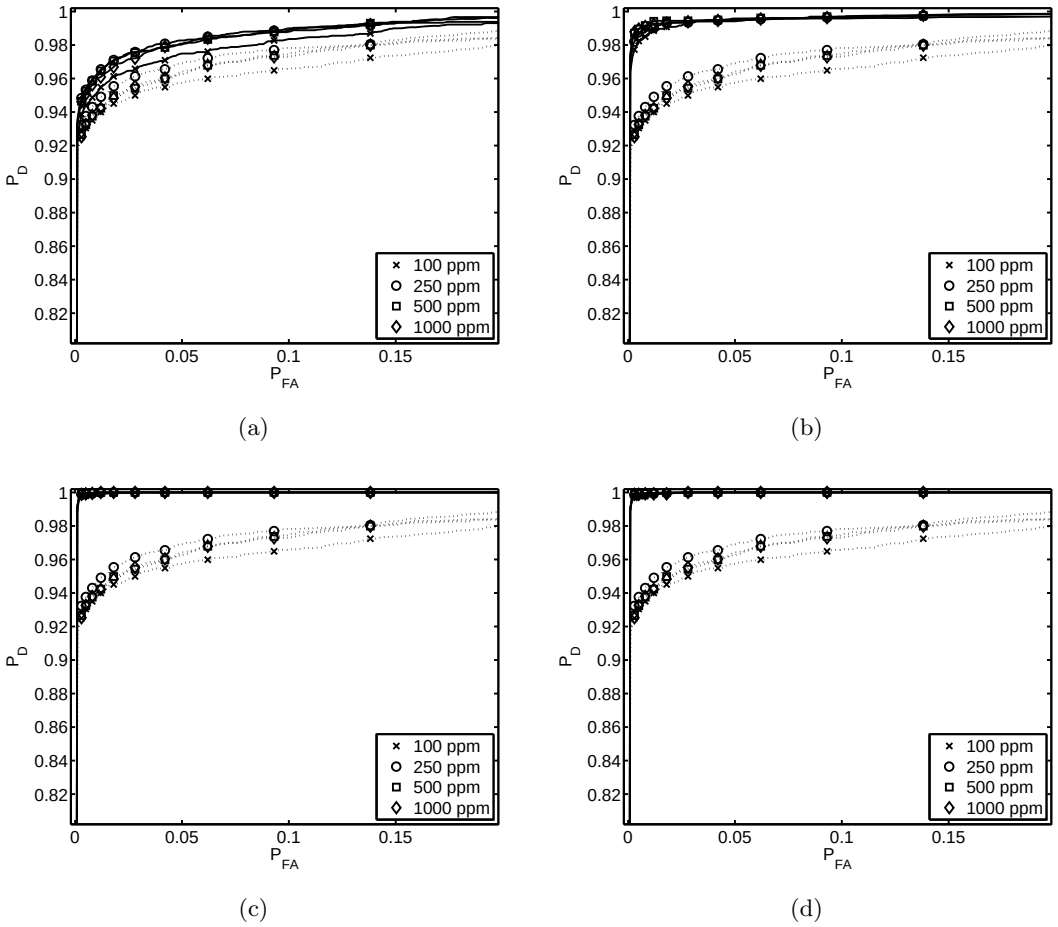


Fig. B.16: B–SIDE filter and LDA at mid-lat winter, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

Table B.3: Area under ROC curve for LDA and B–SIDE + LDA at mid-lat summer.

				$A_{B_{LDA}}$, $\alpha =$			
SNR (dB)	elevation (m)	rate (ppm)	A_{LDA}	1	10	10^2	10^3
50	5k	100	0.99116	0.99429	0.99766	0.99926	0.99926
		250	0.99238	0.99524	0.99747	0.99937	0.99948
		500	0.9928	0.99514	0.99832	0.99948	0.99948
		1000	0.99393	0.99611	0.99863	0.99948	0.99948
	10k	100	0.98262	0.99006	0.99663	0.99907	0.99905
		250	0.98744	0.99266	0.99792	0.99946	0.99946
		500	0.98484	0.99268	0.99851	0.99948	0.99948
		1000	0.98539	0.99374	0.9989	0.99948	0.99948
60	5k	100	0.99481	0.99662	0.99865	0.9993	0.99947
		250	0.99544	0.99694	0.99847	0.99939	0.99948
		500	0.99564	0.99724	0.99891	0.99944	0.99948
		1000	0.99633	0.99725	0.99841	0.99899	0.99947
	10k	100	0.99076	0.99315	0.9963	0.99912	0.99941
		250	0.99092	0.99413	0.99748	0.99939	0.99945
		500	0.99058	0.99327	0.99602	0.99929	0.99945
		1000	0.99121	0.99369	0.99684	0.99938	0.99946

Table B.4: Area under ROC curve for LDA and B–SIDE + LDA at mid-lat winter.

				$A_{B_{LDA}}$, $\alpha =$			
SNR (dB)	elevation (m)	rate (ppm)	A_{LDA}	1	10	10^2	10^3
50	5k	100	0.98667	0.99417	0.99861	0.99915	0.99919
		250	0.98681	0.99439	0.99827	0.9994	0.99948
		500	0.98689	0.99532	0.99896	0.99949	0.99949
		1000	0.98885	0.99634	0.99942	0.99949	0.99949
	10k	100	0.96667	0.98278	0.99668	0.99875	0.99875
		250	0.97099	0.98615	0.99757	0.99923	0.99923
		500	0.9697	0.98851	0.99889	0.99928	0.99928
		1000	0.97299	0.99131	0.9993	0.99939	0.99939
60	5k	100	0.99478	0.99711	0.99885	0.99941	0.99946
		250	0.99525	0.99757	0.99845	0.9993	0.99947
		500	0.9959	0.99747	0.99892	0.99946	0.99948
		1000	0.99643	0.99768	0.99847	0.99906	0.99947
	10k	100	0.98802	0.99431	0.9978	0.99947	0.99946
		250	0.9918	0.99608	0.99845	0.99948	0.99948
		500	0.99042	0.99595	0.99848	0.99949	0.99948
		1000	0.9896	0.9952	0.9984	0.99948	0.99947

B.3 QDA

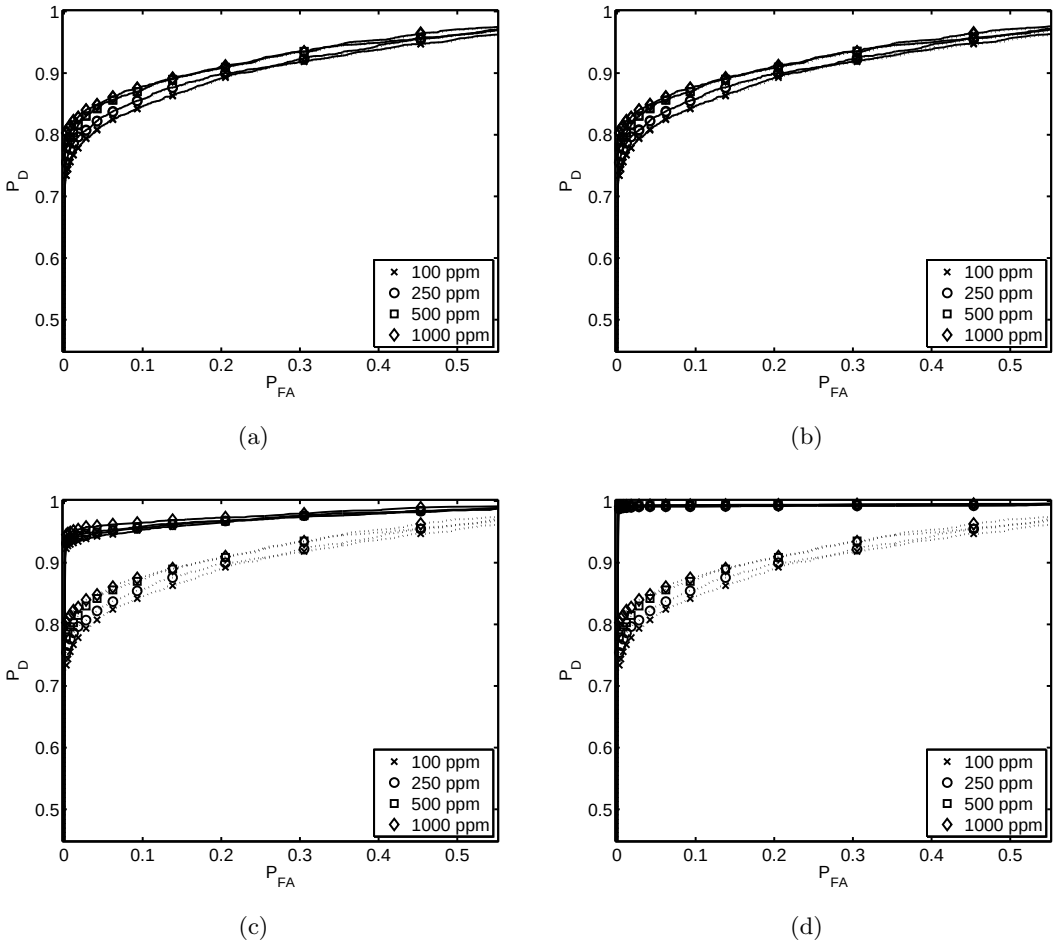


Fig. B.17: B–SIDE filter and QDA at mid-lat summer, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

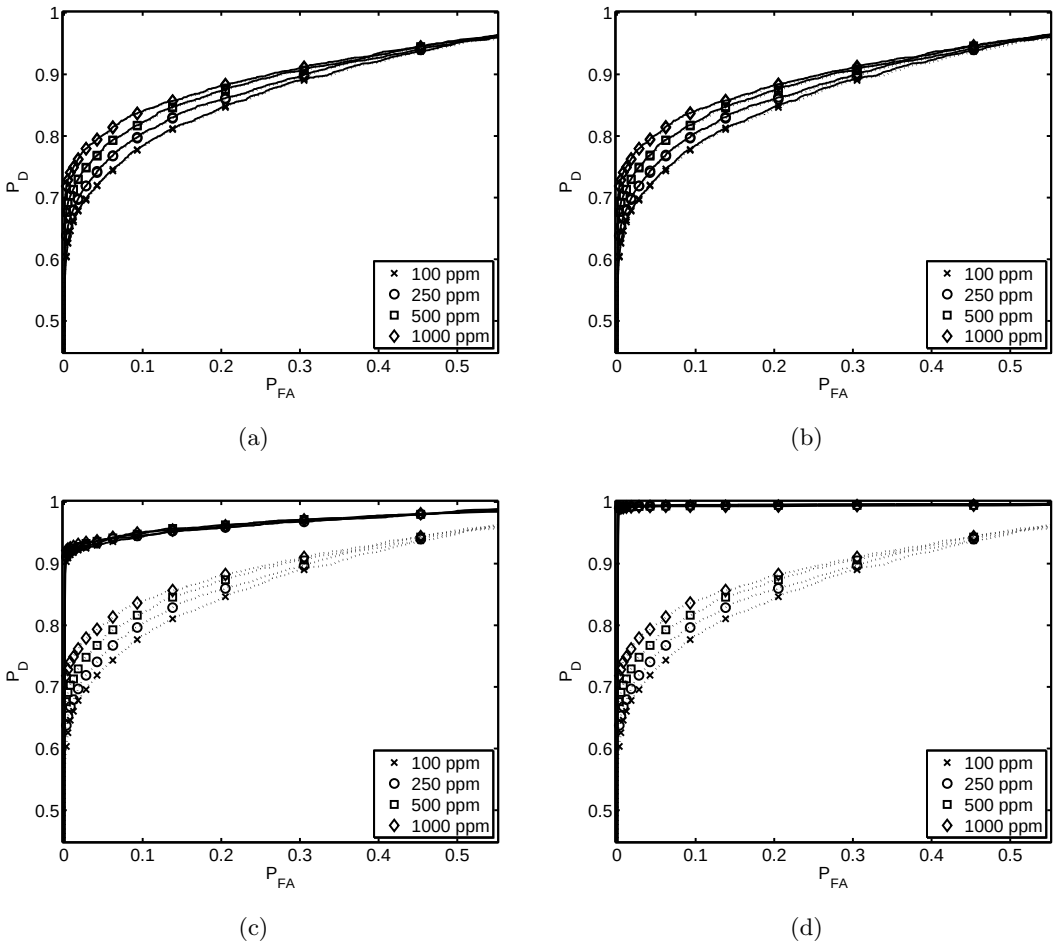


Fig. B.18: B–SIDE filter and QDA at mid-lat summer, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

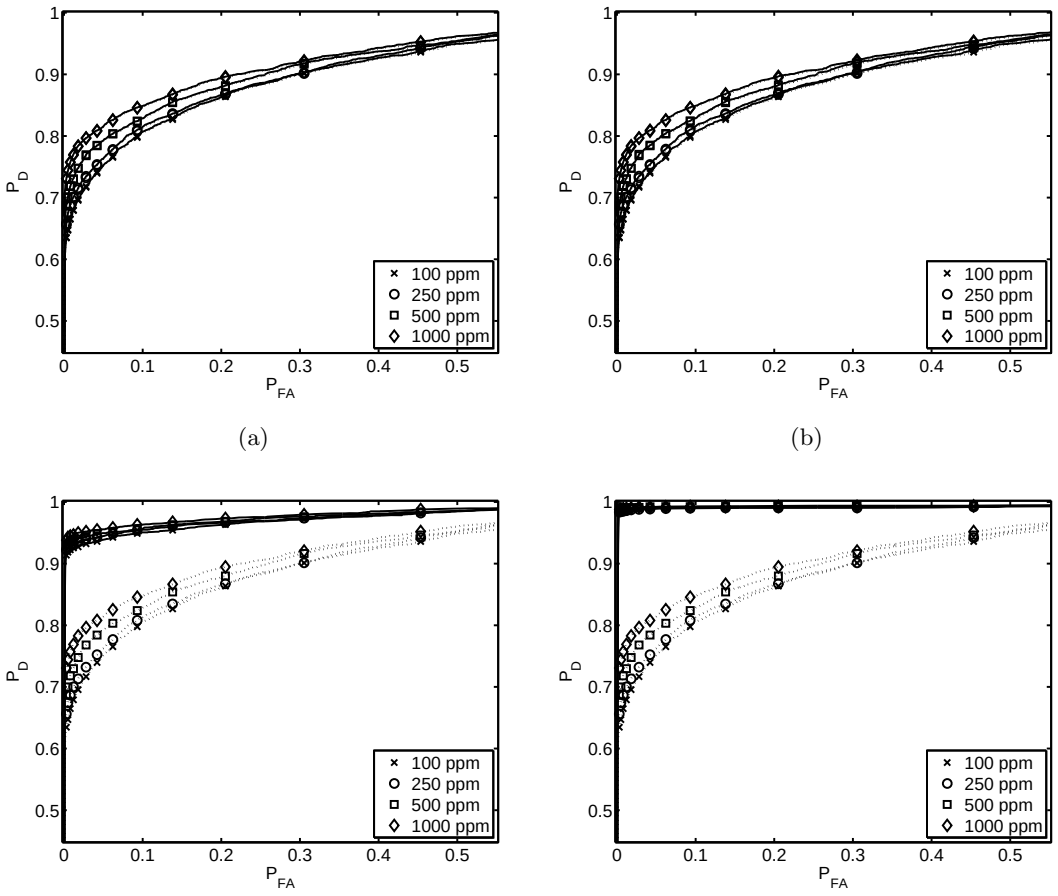


Fig. B.19: B–SIDE filter and QDA at mid-lat winter, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

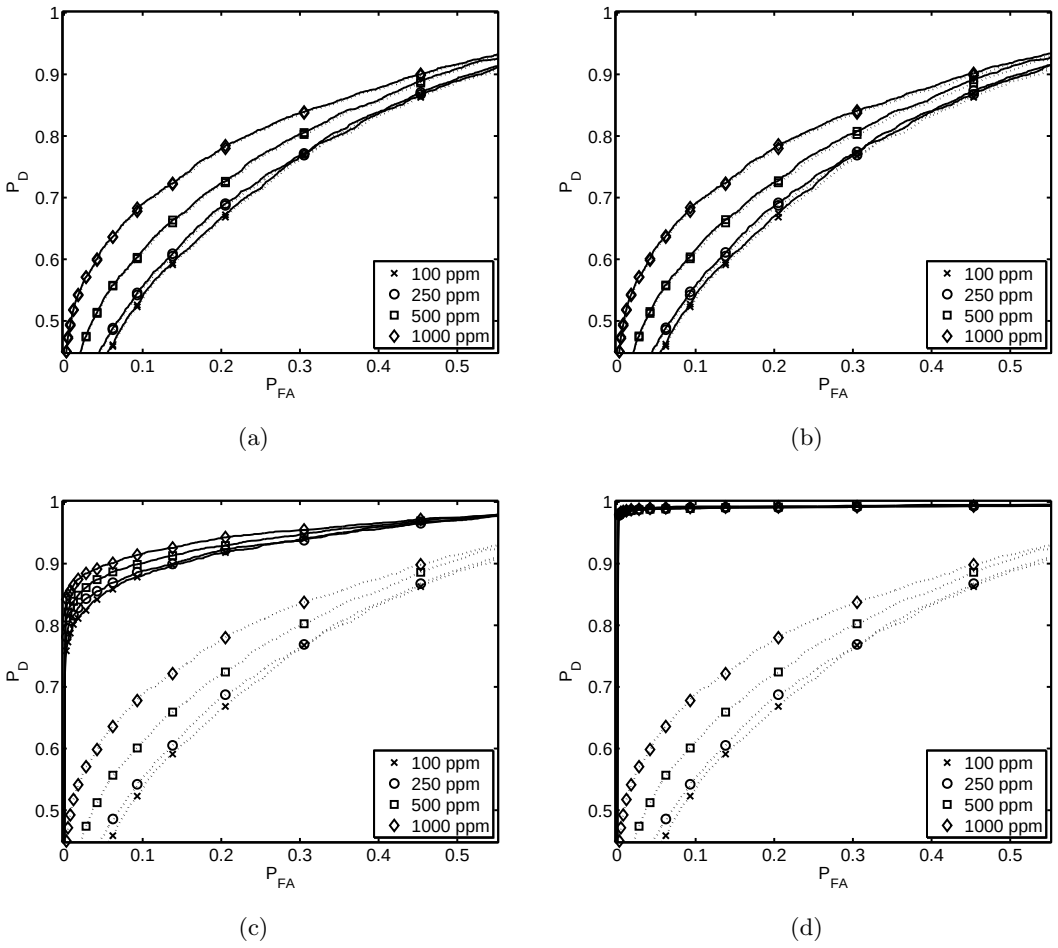


Fig. B.20: B–SIDE filter and QDA at mid-lat winter, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

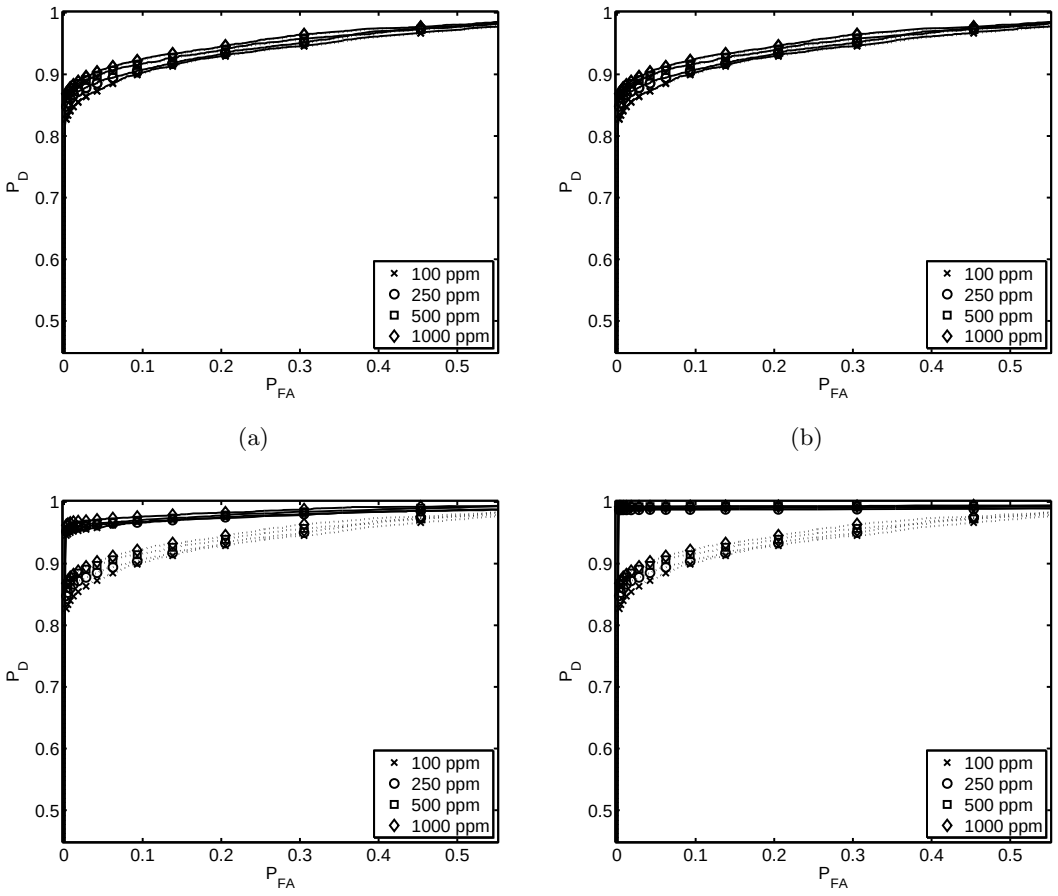


Fig. B.21: B–SIDE filter and QDA at mid-lat summer, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

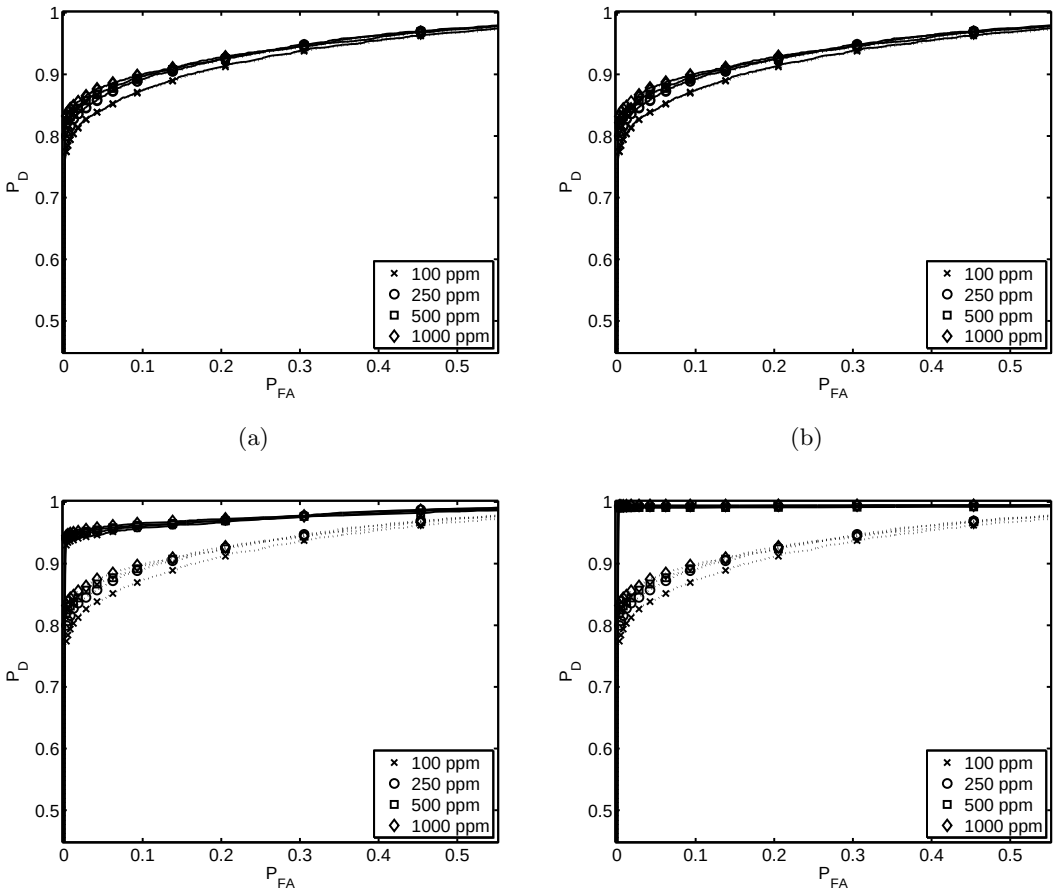


Fig. B.22: B–SIDE filter and QDA at mid-lat summer, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

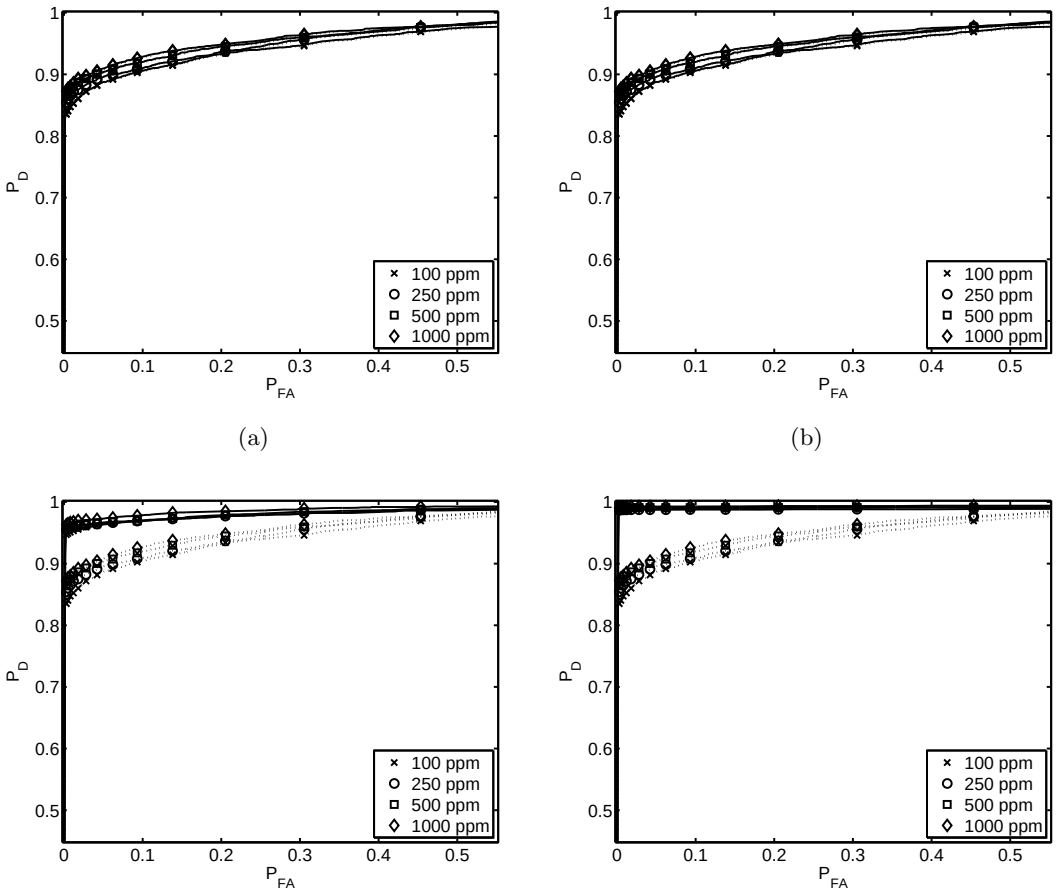


Fig. B.23: B–SIDE filter and QDA at mid-lat winter, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

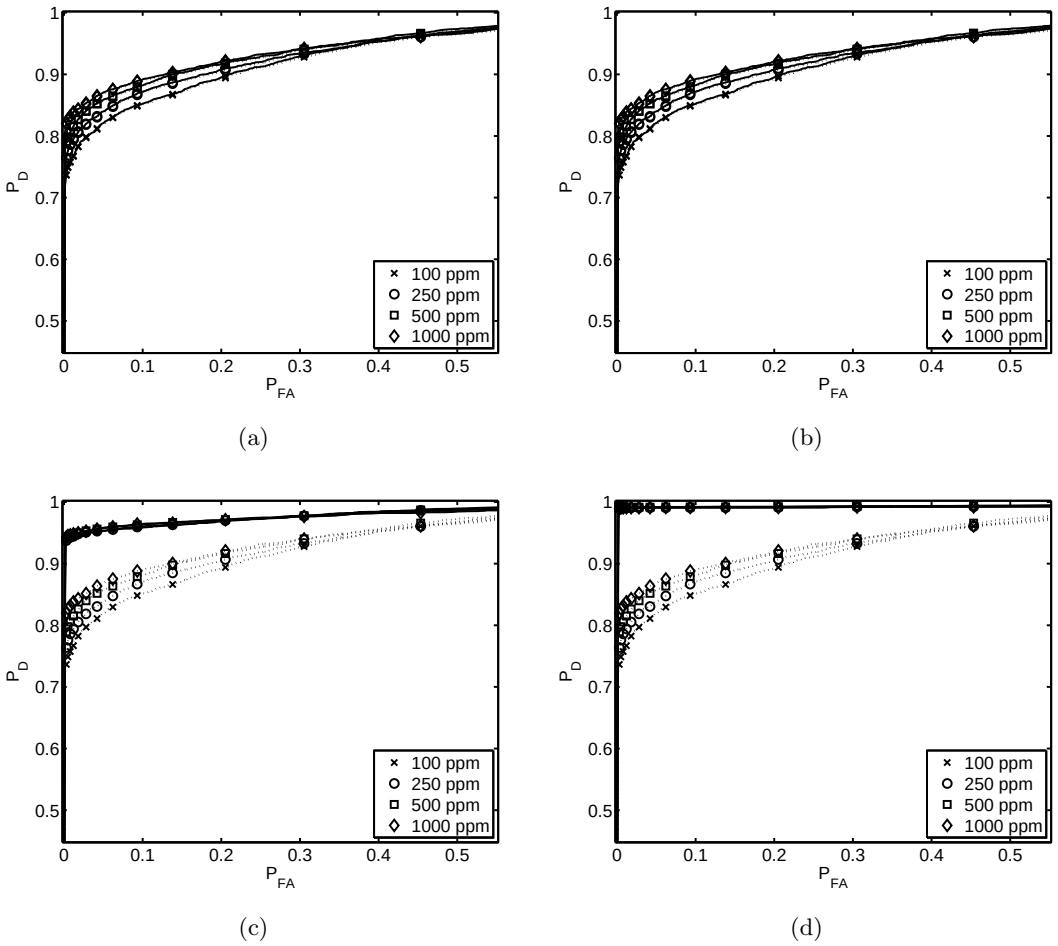


Fig. B.24: B–SIDE filter and QDA at mid-lat winter, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

Table B.5: Area under ROC curve for QDA and B–SIDE + QDA at mid-lat summer.

SNR (dB)	elevation (m)	rate (ppm)	A_{QDA}	$A_{B_{QDA}}$, $\alpha =$ 1	10	10^2	10^3
50	5k	100	0.93703	0.93791	0.93852	0.98022	0.99521
		250	0.94248	0.94328	0.94384	0.98154	0.99421
		500	0.94882	0.94956	0.95004	0.9824	0.99557
		1000	0.95144	0.95217	0.95262	0.98549	0.99578
	10k	100	0.9156	0.91678	0.91767	0.97663	0.99678
		250	0.92216	0.92327	0.92408	0.97725	0.99585
		500	0.92835	0.92937	0.9301	0.97844	0.99551
		1000	0.93225	0.9332	0.93391	0.97783	0.99512
60	5k	100	0.95956	0.96044	0.96061	0.98479	0.99332
		250	0.96353	0.96434	0.96452	0.98485	0.99123
		500	0.96663	0.96739	0.96756	0.98745	0.99417
		1000	0.96977	0.97054	0.97067	0.98964	0.99466
	10k	100	0.94987	0.95085	0.95108	0.98134	0.99392
		250	0.95724	0.95808	0.9583	0.98351	0.99509
		500	0.9578	0.9587	0.95892	0.9819	0.99335
		1000	0.95929	0.96021	0.96024	0.98449	0.99532

Table B.6: Area under ROC curve for QDA and B–SIDE + QDA at mid-lat winter.

SNR (dB)	elevation (m)	rate (ppm)	A_{QDA}	$A_{B_{QDA}}$, $\alpha =$ 1	10	10^2	10^3
50	5k	100	0.92103	0.92218	0.92295	0.97905	0.99431
		250	0.92491	0.92601	0.92677	0.98063	0.99341
		500	0.93314	0.93414	0.9348	0.98198	0.99509
		1000	0.93997	0.94087	0.94146	0.98457	0.99491
	10k	100	0.82027	0.82295	0.82481	0.95414	0.99504
		250	0.8268	0.82941	0.8312	0.95506	0.99357
		500	0.85007	0.85233	0.85387	0.96061	0.99385
		1000	0.87454	0.87643	0.87775	0.9656	0.99382
60	5k	100	0.96085	0.96176	0.9619	0.98533	0.99284
		250	0.96548	0.9663	0.96644	0.9852	0.99079
		500	0.96844	0.96922	0.96934	0.98678	0.99285
		1000	0.97069	0.97145	0.97157	0.98996	0.99439
	10k	100	0.94388	0.94523	0.9454	0.98199	0.99424
		250	0.94836	0.94961	0.9498	0.98249	0.99308
		500	0.95385	0.95495	0.95514	0.98401	0.9941
		1000	0.95429	0.95547	0.95568	0.98238	0.993

B.4 LC

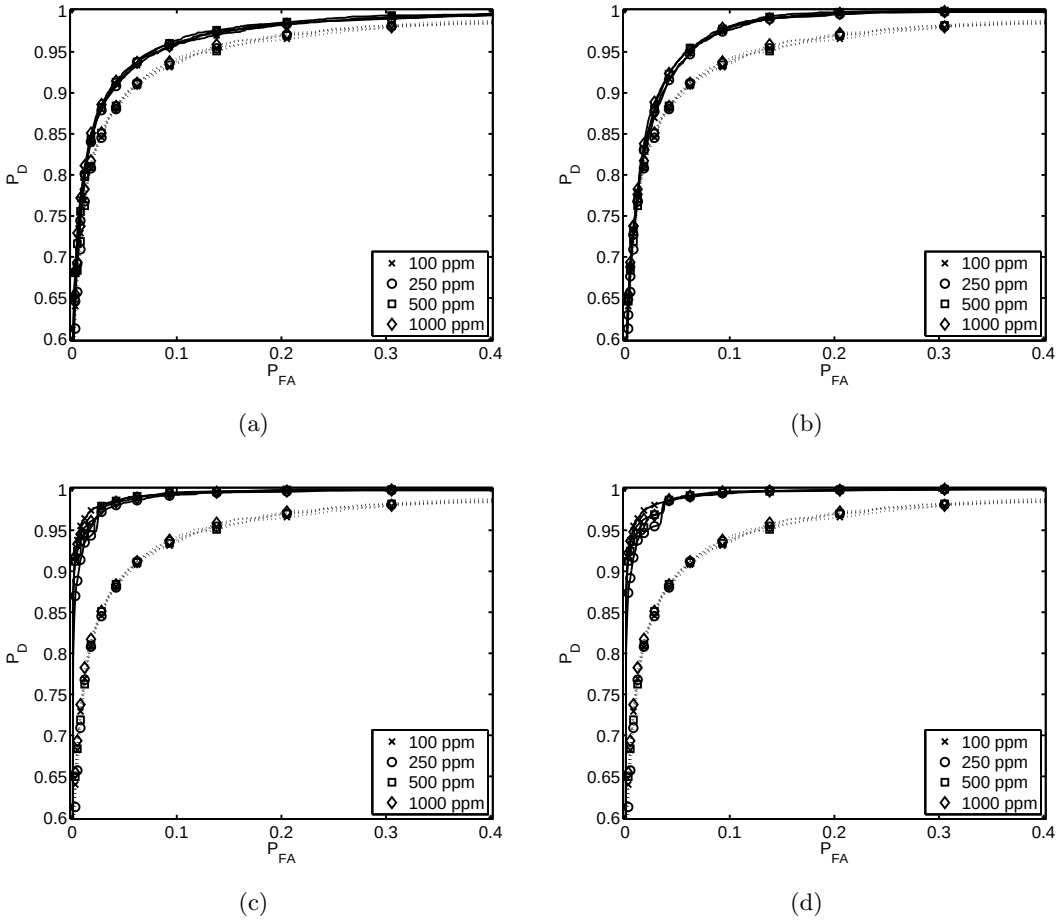


Fig. B.25: B–SIDE filter and LC at mid-lat summer, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

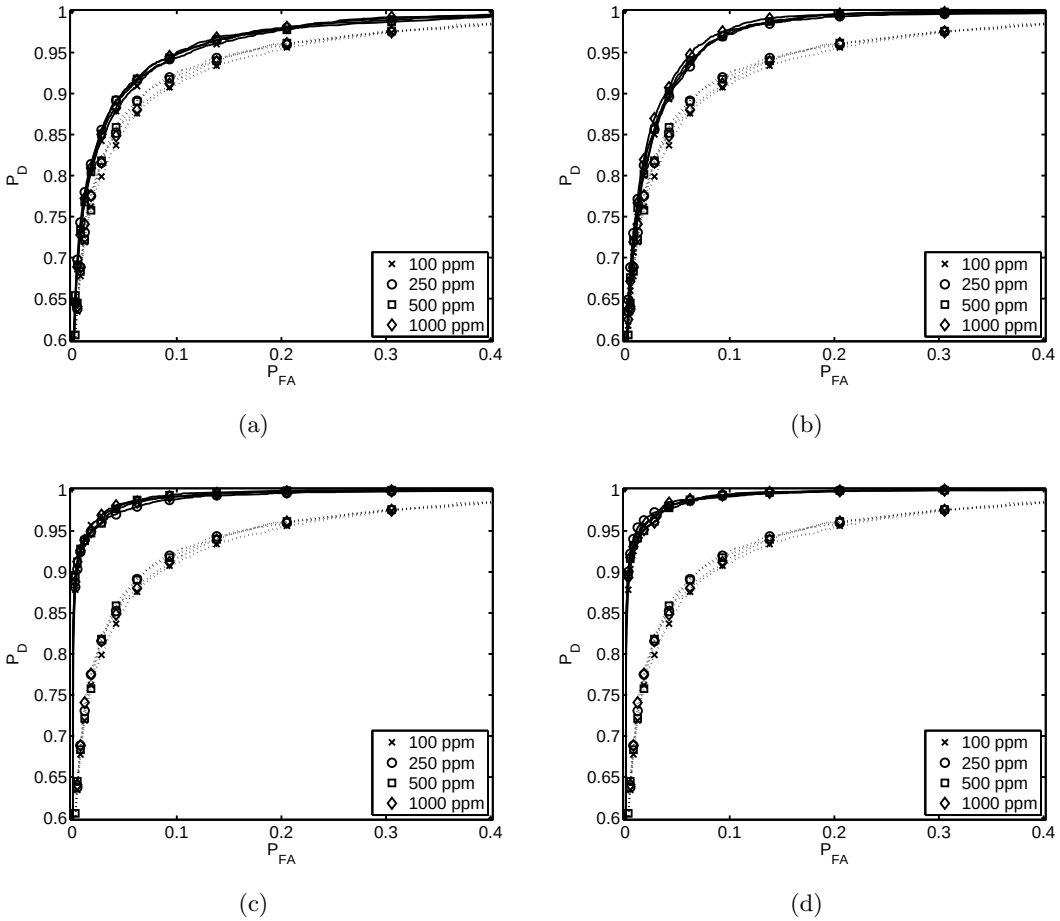


Fig. B.26: B–SIDE filter and LC at mid-lat summer, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

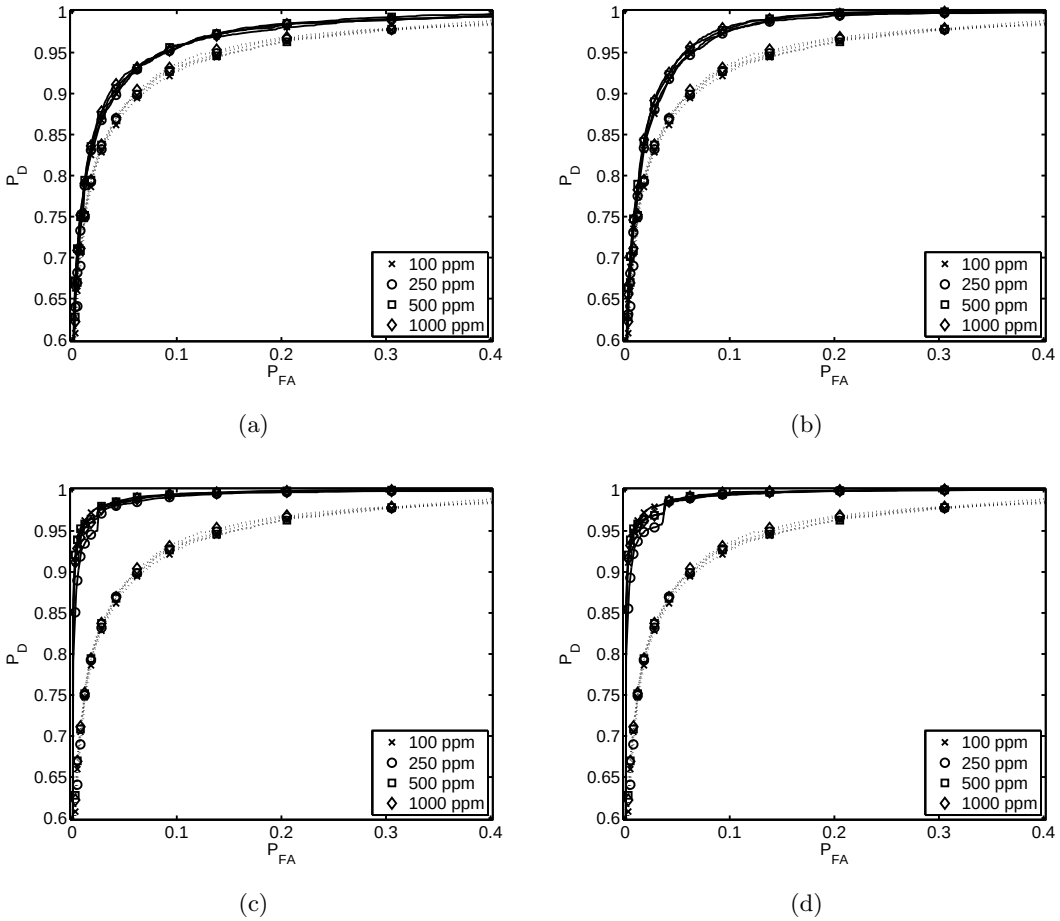


Fig. B.27: B–SIDE filter and LC at mid-lat winter, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

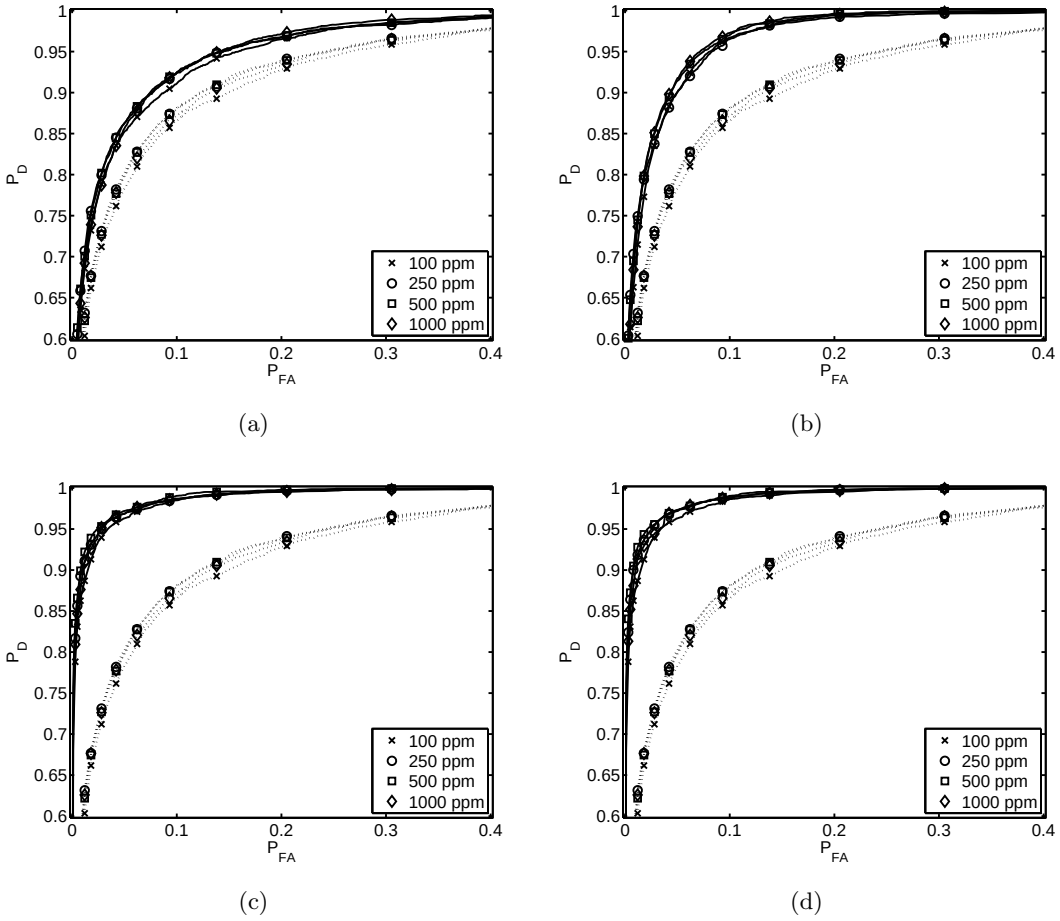


Fig. B.28: B–SIDE filter and LC at mid-lat winter, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

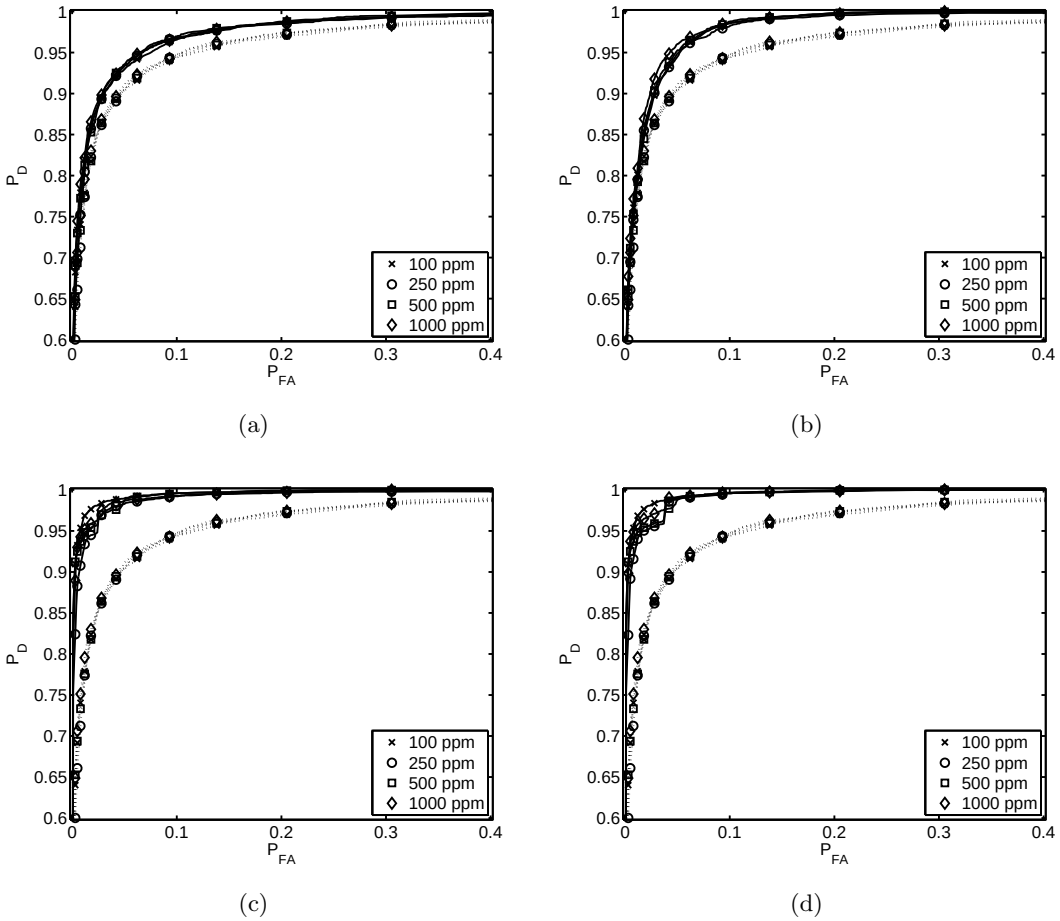


Fig. B.29: B–SIDE filter and LC at mid-lat summer, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

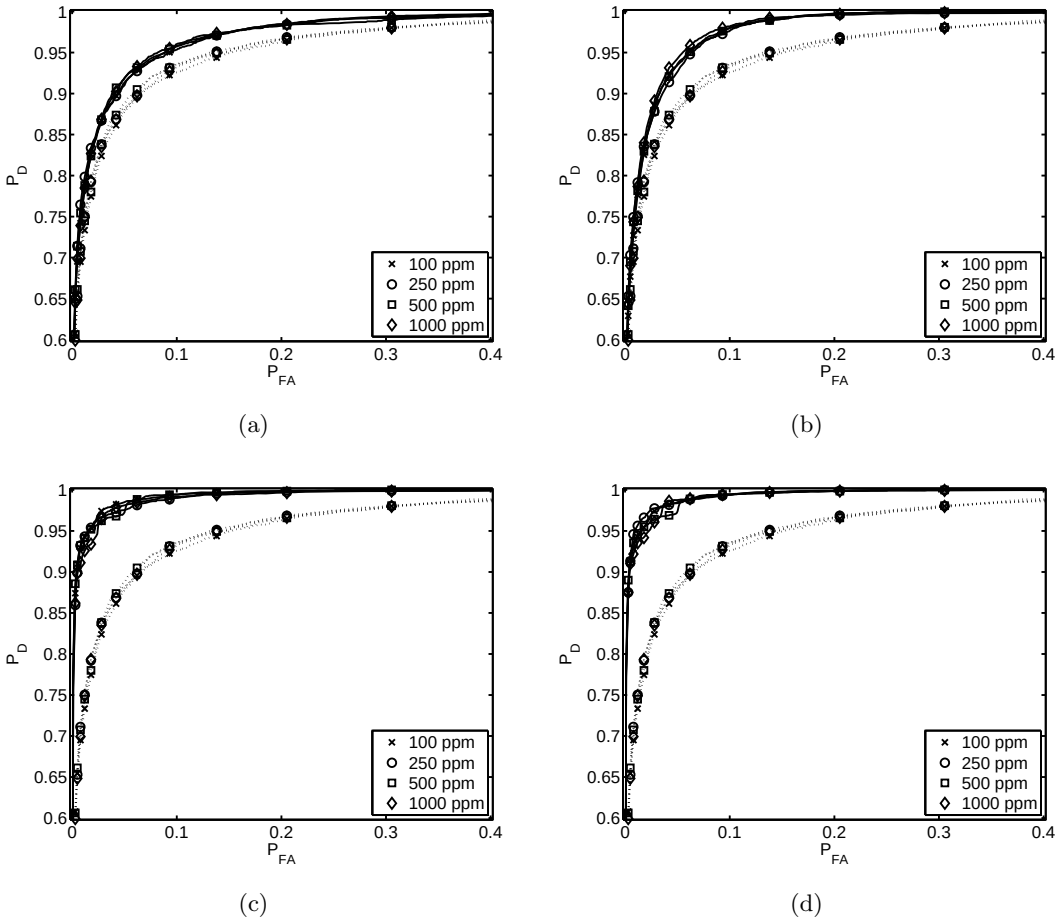


Fig. B.30: B–SIDE filter and LC at mid-lat summer, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

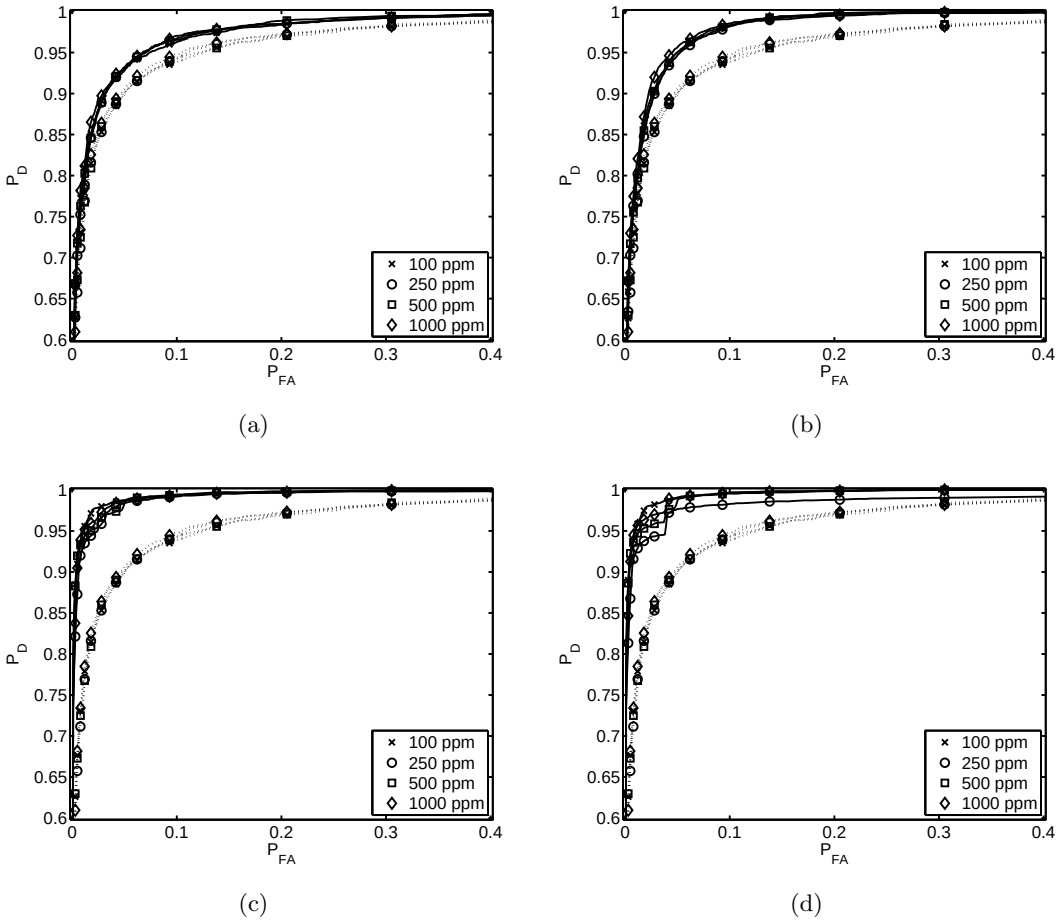


Fig. B.31: B–SIDE filter and LC at mid-lat winter, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

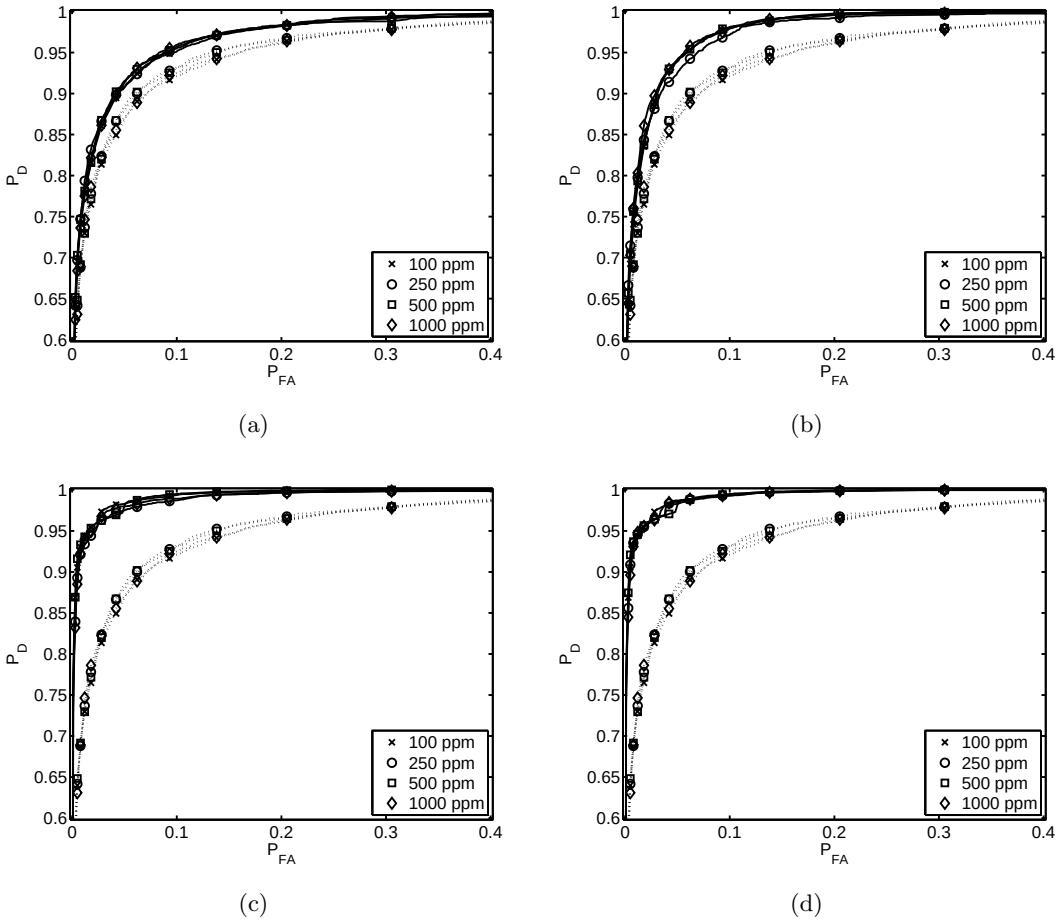


Fig. B.32: B–SIDE filter and LC at mid-lat winter, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

Table B.7: Area under ROC curve for LC and B–SIDE + LC at mid-lat summer.

				$A_{B_{LC}}$, $\alpha =$			
SNR (dB)	elevation (m)	rate (ppm)	A_{LC}	1	10	10^2	10^3
50	5k	100	0.97437	0.98414	0.98816	0.99726	0.99726
		250	0.97517	0.98362	0.98737	0.99533	0.99632
		500	0.97598	0.98527	0.98889	0.99702	0.99702
		1000	0.9753	0.98407	0.98862	0.99667	0.99717
	10k	100	0.96782	0.98017	0.98627	0.99601	0.99601
		250	0.97026	0.98069	0.98581	0.99478	0.99673
		500	0.96977	0.98016	0.98667	0.99621	0.99634
		1000	0.96931	0.98106	0.98755	0.99578	0.99628
60	5k	100	0.97679	0.9861	0.99013	0.99742	0.99742
		250	0.97702	0.98539	0.98886	0.99476	0.99611
		500	0.97827	0.98686	0.99027	0.99671	0.99671
		1000	0.97753	0.98612	0.99088	0.99625	0.99711
	10k	100	0.97262	0.98345	0.98846	0.99638	0.99638
		250	0.97403	0.98303	0.98761	0.99532	0.99673
		500	0.97332	0.98283	0.98848	0.99613	0.99626
		1000	0.97259	0.98346	0.98906	0.99515	0.99615

Table B.8: Area under ROC curve for LC and B–SIDE + LC at mid-lat winter.

				$A_{B_{LC}}$, $\alpha =$			
SNR (dB)	elevation (m)	rate (ppm)	A_{LC}	1	10	10^2	10^3
50	5k	100	0.97093	0.98253	0.98813	0.99697	0.99697
		250	0.97217	0.98214	0.98713	0.99523	0.99615
		500	0.97298	0.98407	0.98923	0.99716	0.99715
		1000	0.973	0.9831	0.98904	0.9966	0.99707
	10k	100	0.95143	0.97113	0.98383	0.99283	0.99283
		250	0.95569	0.9727	0.98325	0.99353	0.99448
		500	0.95604	0.97319	0.98573	0.99474	0.99494
		1000	0.95442	0.97309	0.98545	0.99344	0.99403
60	5k	100	0.97582	0.98529	0.99019	0.99663	0.99711
		250	0.97609	0.9849	0.98879	0.99469	0.98939
		500	0.97694	0.98601	0.98996	0.99612	0.99649
		1000	0.97685	0.98558	0.99061	0.99607	0.9968
	10k	100	0.9714	0.98291	0.98888	0.9963	0.9963
		250	0.97278	0.98233	0.98686	0.99458	0.99639
		500	0.97226	0.98262	0.98903	0.99617	0.99631
		1000	0.97089	0.98285	0.98911	0.99521	0.99617

B.5 Fisher's Discriminant

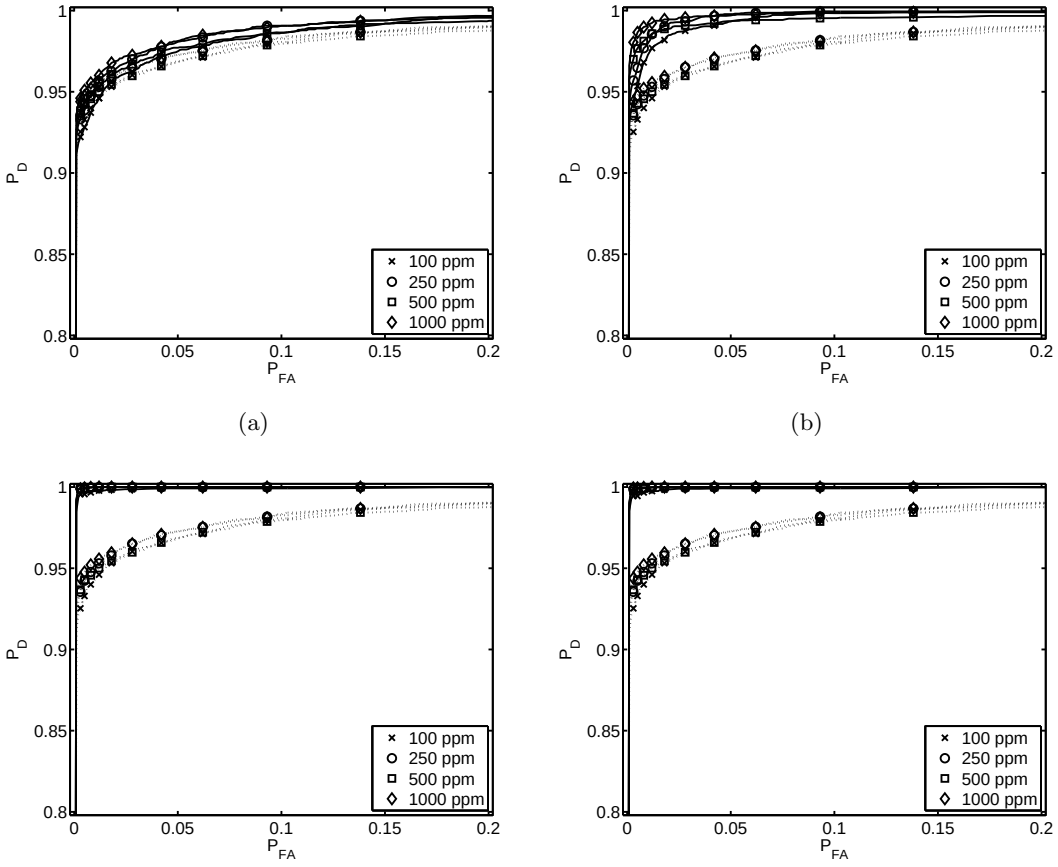


Fig. B.33: B–SIDE filter and Fisher's discriminant at mid-lat summer, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

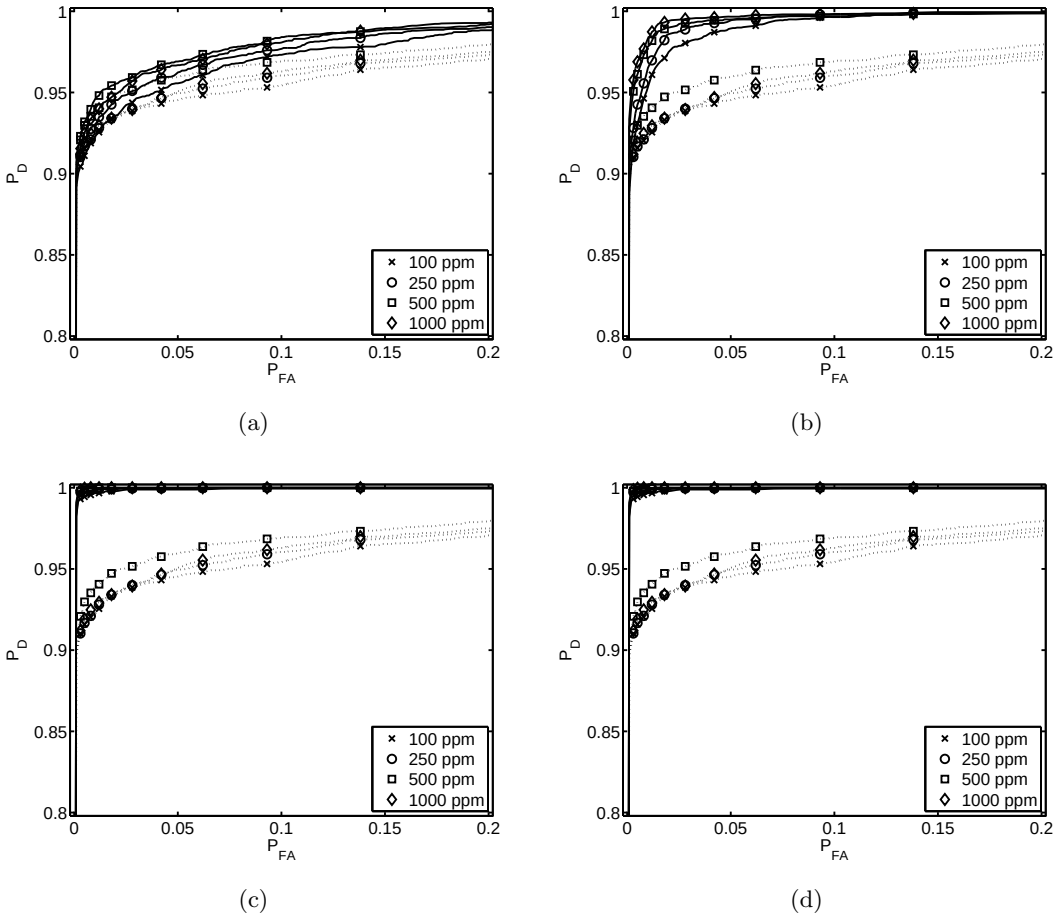


Fig. B.34: B–SIDE filter and Fisher's discriminant at mid-lat summer, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

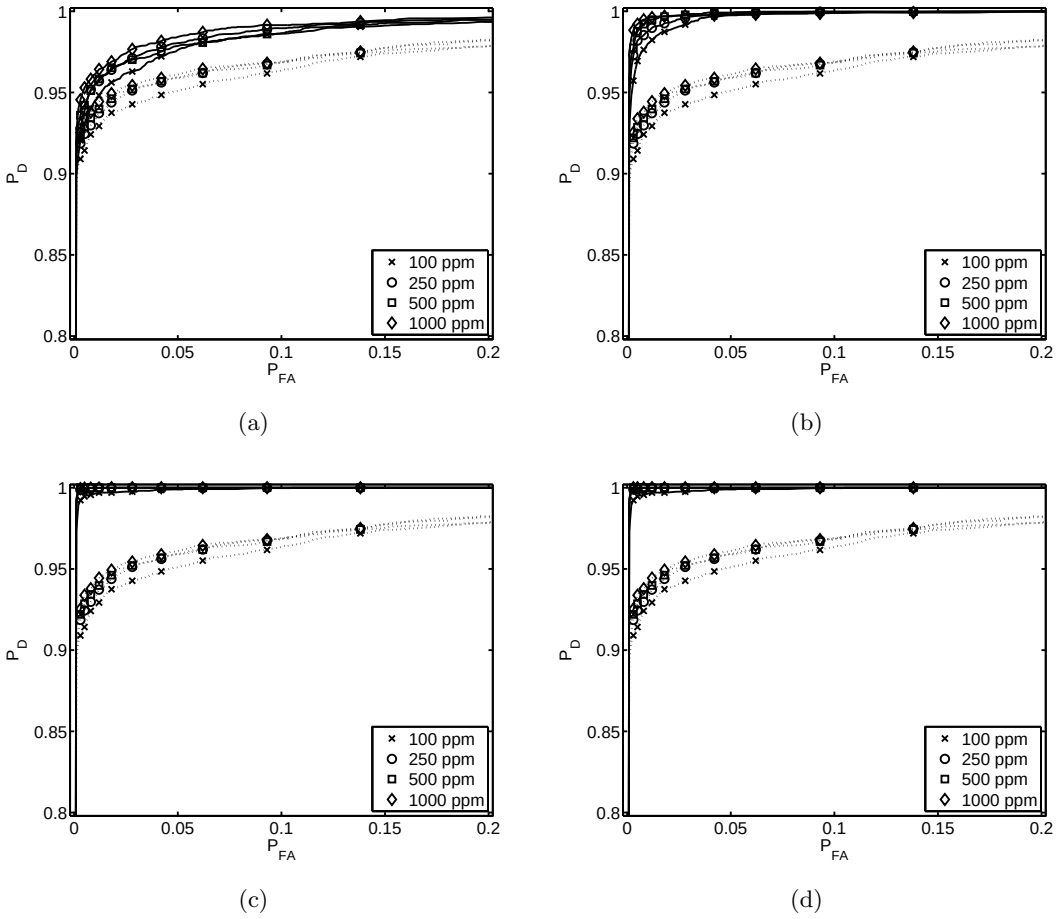


Fig. B.35: B–SIDE filter and Fisher's discriminant at mid-lat winter, $SNR = 50$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

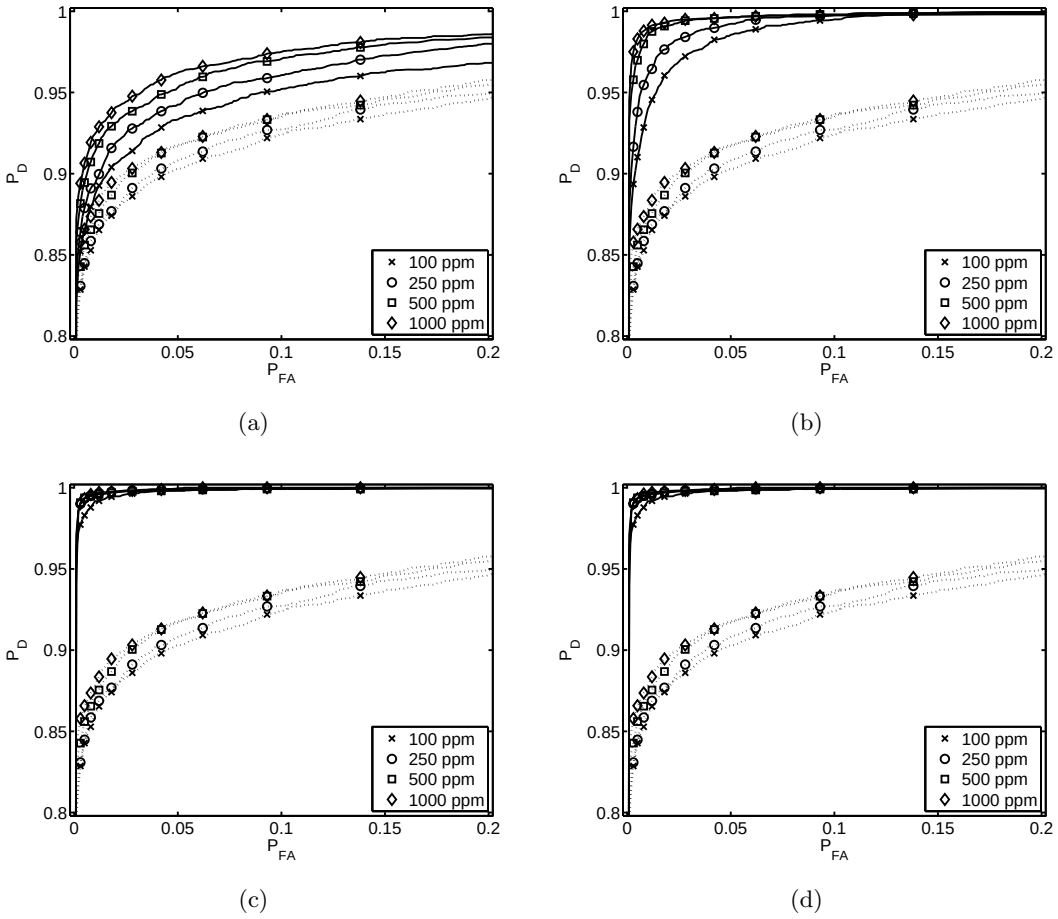


Fig. B.36: B–SIDE filter and Fisher's discriminant at mid-lat winter, $SNR = 50$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

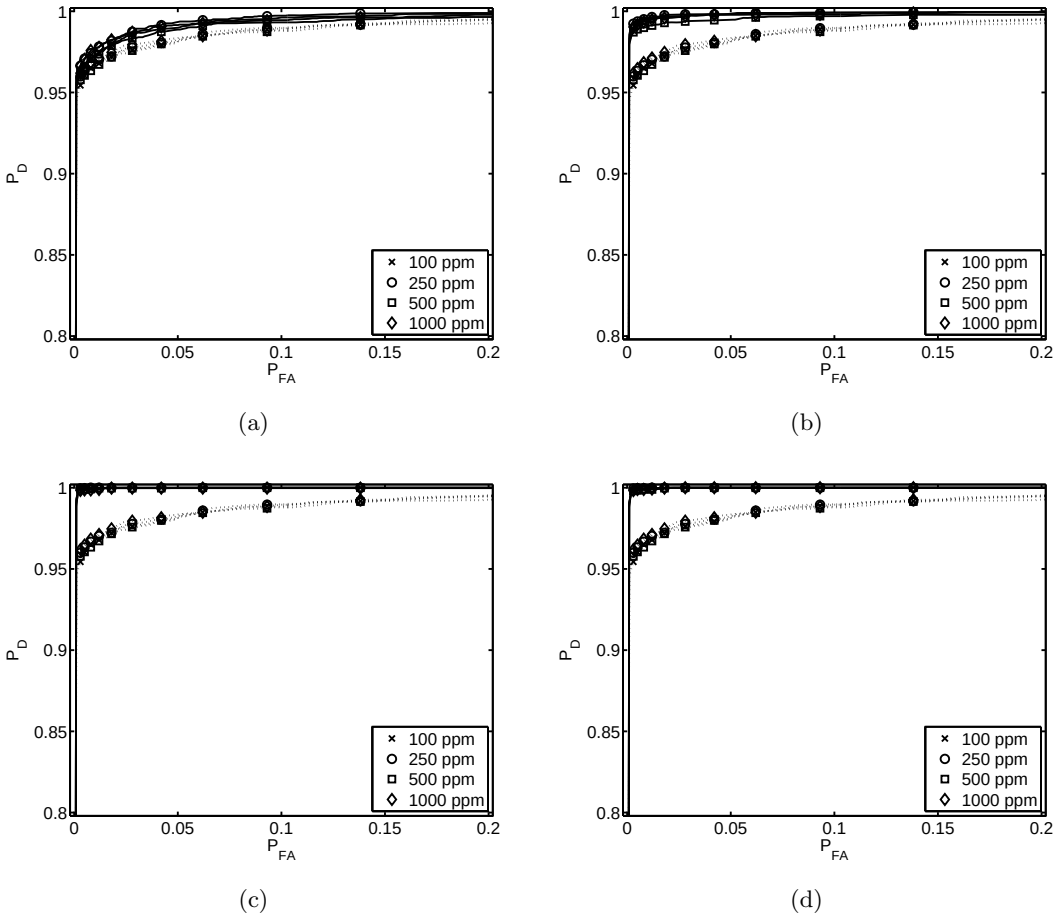


Fig. B.37: B–SIDE filter and Fisher's discriminant at mid-lat summer, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

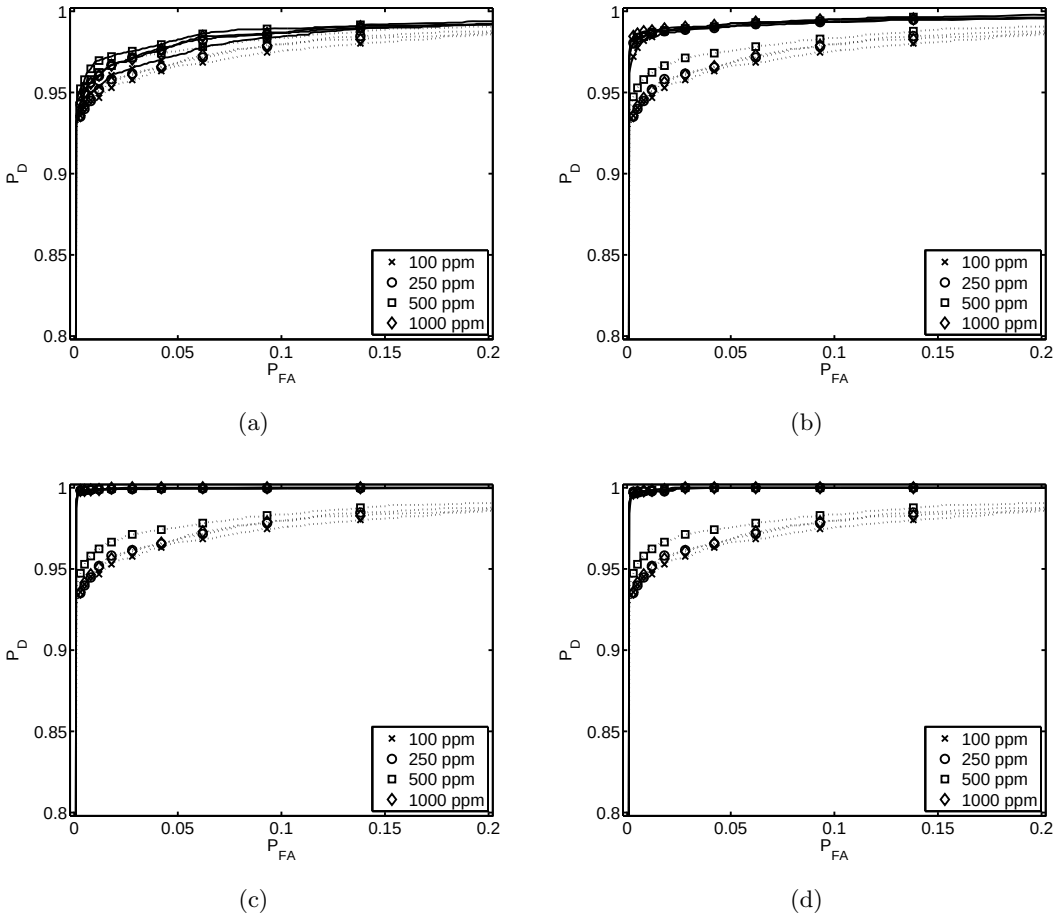


Fig. B.38: B–SIDE filter and Fisher's discriminant at mid-lat summer, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

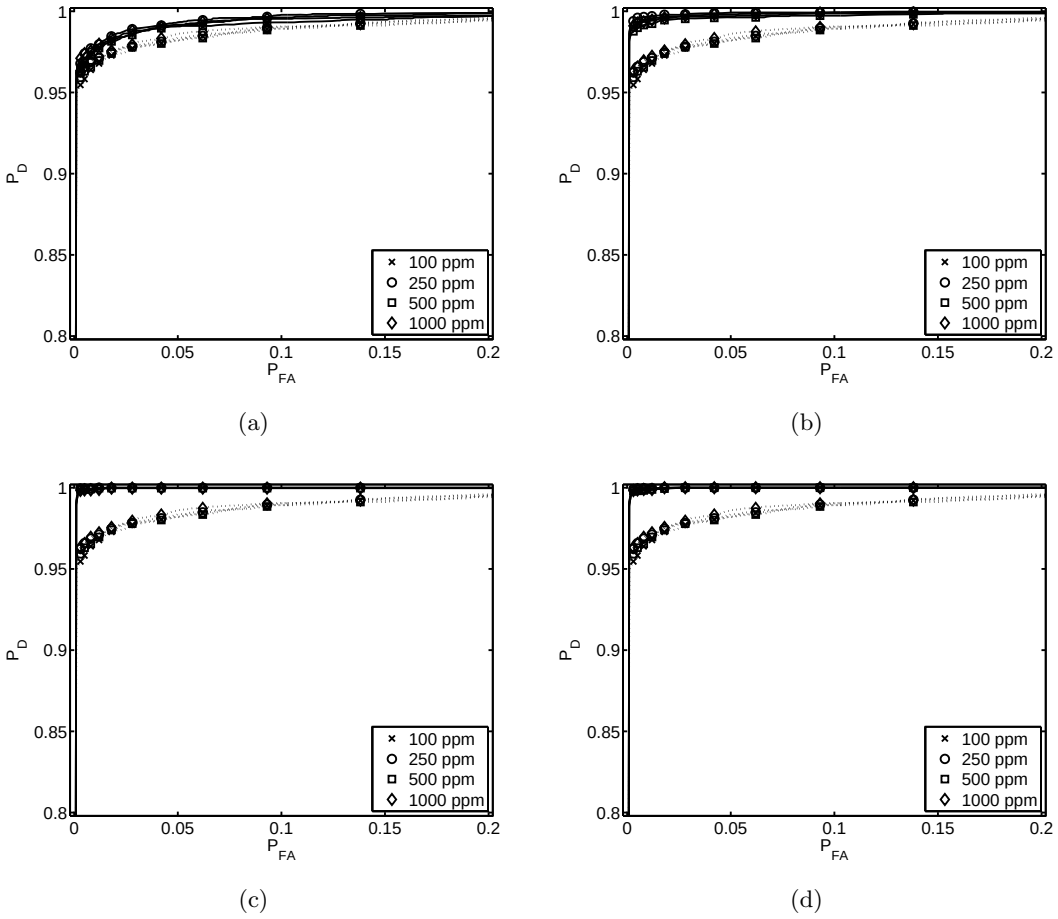


Fig. B.39: B–SIDE filter and Fisher's discriminant at mid-lat winter, $SNR = 60$ dB, sensor at 5 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

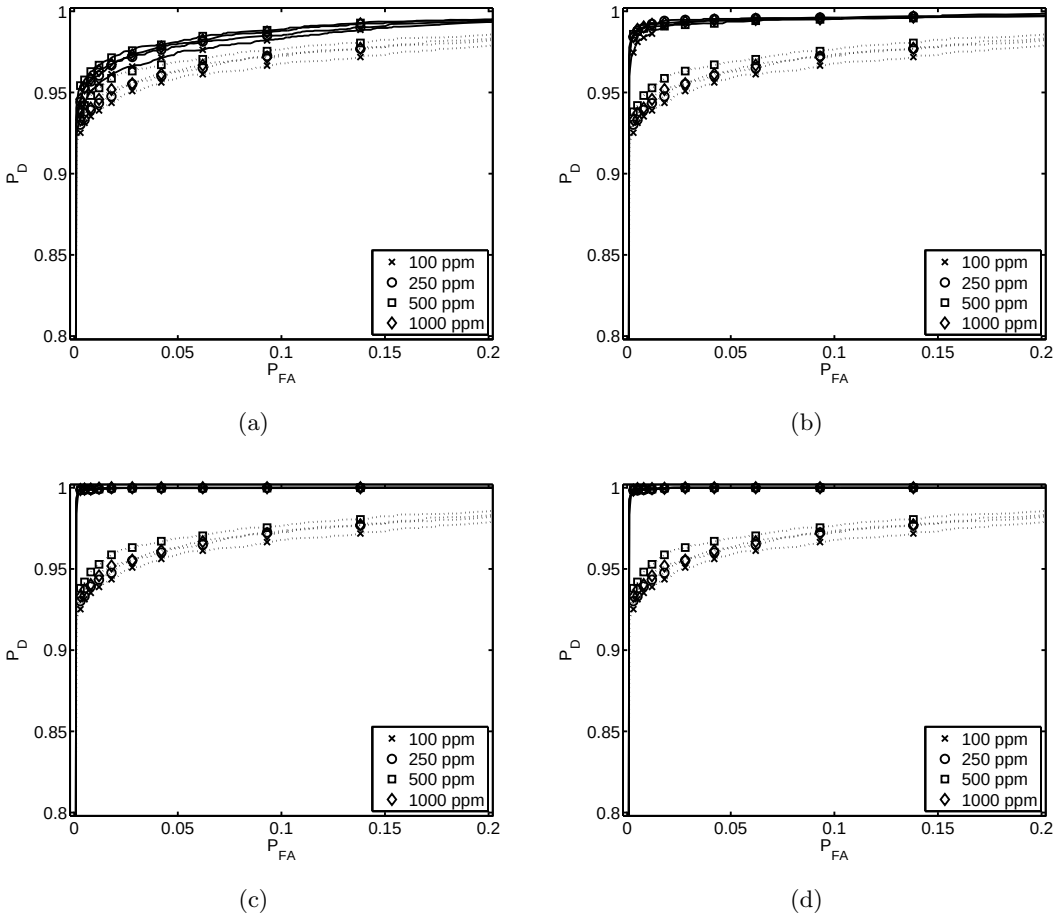


Fig. B.40: B–SIDE filter and Fisher's discriminant at mid-lat winter, $SNR = 60$ dB, sensor at 10 km. In (a) $\alpha = 1$, (b) $\alpha = 10$, (c) $\alpha = 100$, and (d) $\alpha = 1000$.

Table B.9: Area under ROC curve for Fisher's discriminant and B–SIDE + Fisher's discriminant at mid-lat summer.

				$A_{B_{FISH}}$, $\alpha =$			
SNR (dB)	elevation (m)	rate (ppm)	A_{FISH}	1	10	10^2	10^3
50	5k	100	0.99268	0.99523	0.99824	0.9993	0.9993
		250	0.99325	0.99586	0.99847	0.99949	0.99948
		500	0.99129	0.99375	0.99692	0.99948	0.99948
		1000	0.99336	0.99582	0.99856	0.99948	0.99948
	10k	100	0.98312	0.99102	0.99736	0.99921	0.99921
		250	0.98473	0.99218	0.9981	0.99944	0.99944
		500	0.98713	0.99335	0.99823	0.99948	0.99948
		1000	0.98446	0.99343	0.99826	0.99948	0.99948
60	5k	100	0.99577	0.99769	0.99912	0.99948	0.99948
		250	0.99595	0.99813	0.99917	0.99949	0.99949
		500	0.99489	0.99665	0.99818	0.99947	0.99947
		1000	0.996	0.99753	0.99892	0.9994	0.99947
	10k	100	0.99023	0.9933	0.99661	0.99924	0.99942
		250	0.99113	0.99359	0.99694	0.99931	0.99944
		500	0.99246	0.99452	0.9975	0.99945	0.99946
		1000	0.99126	0.99421	0.99707	0.99946	0.99944

Table B.10: Area under ROC curve for Fisher's discriminant and B–SIDE + Fisher's discriminant at mid-lat winter.

				$A_{B_{FISH}}$, $\alpha =$			
SNR (dB)	elevation (m)	rate (ppm)	A_{FISH}	1	10	10^2	10^3
50	5k	100	0.98673	0.99473	0.99871	0.9993	0.9993
		250	0.98842	0.99546	0.99904	0.99949	0.99949
		500	0.98731	0.99522	0.99919	0.99949	0.99949
		1000	0.98846	0.99634	0.99915	0.99948	0.99949
	10k	100	0.96734	0.98255	0.99673	0.99912	0.99912
		250	0.97164	0.98681	0.99784	0.99927	0.99927
		500	0.97361	0.98891	0.99832	0.99906	0.99906
		1000	0.97288	0.99081	0.99812	0.99931	0.99932
60	5k	100	0.99594	0.998	0.99902	0.99947	0.99946
		250	0.99564	0.99808	0.9993	0.99948	0.99948
		500	0.99525	0.99709	0.99861	0.99939	0.99947
		1000	0.99587	0.99737	0.99881	0.99931	0.99947
	10k	100	0.98771	0.99429	0.99789	0.99948	0.99947
		250	0.98948	0.99525	0.99836	0.99945	0.99947
		500	0.99137	0.99605	0.99796	0.99944	0.99948
		1000	0.98921	0.99537	0.99784	0.99947	0.99947